

NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

RELATED APPLICATIONS

This application claims priority from U.S.S.N. 60/242,485 (21402-175), filed October
5 23, 2000; U.S.S.N. 60/263,339 (21402-175A), filed January 22, 2001; U.S.S.N. 60/264,850
(21402-175B), filed January 29, 2001 and Express Mail No.: EL831679474US mailed
October 22, 2001, each of which is incorporated by reference in its entirety.

FIELD OF THE INVENTION

10 The invention generally relates to nucleic acids and polypeptides encoded thereby.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom.
More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear,
15 membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and
recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding
20 novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as NOVX,
or NOV1 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as
derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively
designated as "NOVX" nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated NOVX nucleic acid molecule
25 encoding a NOVX polypeptide that includes a nucleic acid sequence that has identity to the
nucleic acids disclosed in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45,
47, 49, 51, 53, and 55. In some embodiments, the NOVX nucleic acid molecule will hybridize
under stringent conditions to a nucleic acid sequence complementary to a nucleic acid
molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The
30 invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a

fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes
5 the nucleic acid sequence of any of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55) or a complement of said
10 oligonucleotide.

Also included in the invention are substantially purified NOVX polypeptides (SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56). In certain embodiments, the NOVX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

15 The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide,
20 or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression
25 of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex
30 between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

Also included in the invention is a method of detecting the presence of a NOVX nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

5 In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon
10 containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, Cancer, Leukodystrophies, Breast cancer, Ovarian cancer, Prostate cancer, Uterine cancer, Hodgkin disease, Adenocarcinoma, Adrenoleukodystrophy, Cystitis, incontinence, Von Hippel-Lindau
15 (VHL) syndrome, hypercalcaemia, Endometriosis, Hirschsprung's disease, Crohn's Disease, Appendicitis, Cirrhosis, Liver failure, Wolfram Syndrome, Smith-Lemli-Opitz syndrome, Retinitis pigmentosa, Leigh syndrome; Congenital Adrenal Hyperplasia, Xerostomia; tooth decay and other dental problems; Inflammatory bowel disease, Diverticular disease, fertility, Infertility, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic
20 stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, obesity, Diabetes Insipidus and Mellitus with Optic Atrophy and Deafness, Pancreatitis, Metabolic Dysregulation, transplantation recovery, Autoimmune disease,
25 Systemic lupus erythematosus, asthma, arthritis, psoriasis, Emphysema, Scleroderma, allergy, ARDS, Immunodeficiencies, Graft versus host, Alzheimer's disease, Stroke, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Multiple sclerosis, Ataxia-telangiectasia, Behavioral disorders, Addiction, Anxiety, Pain, Neurodegeneration, Muscular dystrophy, Lesch-Nyhan syndrome, Myasthenia gravis, schizophrenia, and other dopamine-dysfunctional states, levodopa-induced dyskinesias, alcoholism, epileptic seizures and other
30 neurological disorders, mental depression, Cerebellar ataxia, pure; Episodic ataxia, type 2; Hemiplegic migraine, Spinocerebellar ataxia-6, Tuberous sclerosis, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Renal tubular acidosis, IgA nephropathy, and/or other pathologies and disorders of the like.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in gene therapy, and NOVX may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

The invention further includes a method for screening for a modulator of disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The method includes contacting a test compound with a NOVX polypeptide and determining if the test compound binds to said NOVX polypeptide. Binding of the test compound to the NOVX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a NOVX nucleic acid. Expression or activity of NOVX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal which recombinantly-expresses NOVX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of NOVX polypeptide in both the test animal and the control animal is compared. A change in the activity of NOVX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide, a NOVX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the amount of the NOVX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the NOVX polypeptide present in a

control sample. An alteration in the level of the NOVX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a NOVX polypeptide, a NOVX nucleic acid, or a NOVX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identify the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to

any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)
1	20936375_0_228_da1	1	2
2a	CG51523-05	21	22
2b	CG51523-05_164786042	23	24
2c	CG51523-05_164732479	25	26
2d	CG51523-05_164732506	27	28
2e	CG51523-05_164732693	29	30
2f	CG51523-05_164732709	31	32
2g	CG51523-05_164718189	33	34
2h	CG51523-05_164718193	35	36
2i	CG51523-05_164718197	37	38
2j	CG51523-05_164718205	39	40
2k	CG51523-05_164718209	41	42
2l	CG51523-05_164718213	43	44
2m	CG51523-05_166190452	45	46
2n	CG51523-05_166190467	47	48
2o	CG51523-05_166190475	49	50
2p	CG51523-05_166190498	51	52
2q	CG51523-05_166190460	53	54
2r	CG51523-05_166190483	55	56

5 NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

10 NOV1 and NOV2 are homologous to an endozepine-related protein precursor-like family of proteins. Thus, the NOV1 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, 15 tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies,

behavioral disorders, addiction, anxiety, pain, neuroprotection, anxiety, depression, stress, immune dysfunction, alcoholism, obesity and diabetes and/or other pathologies/disorders.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, neurogenesis, cell differentiation, cell proliferation, hematopoiesis, wound healing and angiogenesis.

Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOV1

NOV1 includes a novel endozepine-related protein precursor-like protein disclosed below. A disclosed NOV1 nucleic acid of 1747 nucleotides (also referred to as 20936375_0_228_da1) encoding a novel endozepine-related protein precursor-like protein is shown in Table 1A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 38-40 and ending with a TGA codon at nucleotides 1607-1609. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 1A. The start and stop codons are in bold letters.

Table 1A. NOV1 nucleotide sequence (SEQ ID NO:1).

TCTCCTTTTGGGGCATGTTGATCCGCGCTGCGCTCCA**ATG**TTCAGTTTCATGCAGGCTCTTGGGAAAGC
TGGTGCCTGCTGCTGCTGATTCCCGCCGACAGACCTTGGGACCGGGCCAACTGGCAGCTGGAGATGG
CGGACACGAGATCCGTGCACGAGACTAGGTTTGAGGCGGCGTGAAGGTGATCCAGAGTTTGCCGAAGAA
TGGTTTCATTCCAGCCCAACAAATGAAATGATGCTTAAATTTATAGCTTCTATAAGCAGGCAACTGAAGGA
CCCTGTAAACTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGTTTAC
TGGGTGATATGACCAAGAGGAAGCCATGATTGCATATGTTGAAGAAATGAAAAAGATTATTGAACTAT
GCCAATGACTGAGAAAGTTGAAGAATTGCTGCGTGTCTAGGTCCATTTATGAAATTGTCGAGGACAAA
AAGAGTGGCAGGAGTTCTGATATAACCTCAGATCTTGGTAATGTTCTCACTTCTGCTCCGAACGCCAAAA
CCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGT
GAAAGGAGCAGAACAAAGTGATAATGATAAGAAAATGATGAAGAAGTCAGCAGACCATAAGAATTTGGAA
GTCATGTCTACTAATGGCTATGATAAAGATGGCTTTGTTTCAAGATATACAGAATGACATTTCATGCCAGTT
CTTCCTCGAATGGCAGAAGCACTGAAGAAGTAAAGCCATTGATGAAAACTTGGGGCAAACTGGAAAATC
TGCTGTTTGCATTACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTCAGCATTTGACA
AGCGATTTCAGACAGTGAAGTTTACTGTGATTCTATGGAACAATTTGGACAAGAAGAGTCTTTAGACAGCT
TTACGTCCAACAATGGACCATTTTCAGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAAATCTGG
ATTTCTGTAAGATATTCAAGTACCTCCTGGAAATGGCAACATTGGGAATATGCAGGTGGTTGCAGTTGAA
GGAAAAGGTGAAGTCAAGCATGGAGGAGAAGATGGCAGGAATAACAGCGGAGCACCACACCGGGAGAAGC
GAGGCGGAGAACTGACGAATTTCTAATGTTAGAAGAGGAAGAGGACATAGGATACACACTTGAGCGA
AGGAACCAAGGGCCGGCAGGTGGGAAGTGGAGTGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCGA
GGCAGCCTCAATGAGCAGATCGCCCTCGTGTCTGATGAGACTGCAGGAGGACATGCAGAATGTCTTCAGA
GACTGCAGAACTGGAAACGCTGACTGCTTGCAGGCAAAATCATCAACATCAACATTGCAGACTGCTCC
TCAGCCACCTCACAGAGACCATCTTGGTGGCCCTTCGAGATGTCTCCTGGTGTGCTAACGTTTGCCATC
ATATGGCCTTTTATTGCACAGTGGTTGGTGTATTTATACTATCAAAGAAGGAGAAGAAAACTGAAC**TGAG**
GAAAAATGGTGTTTTCTCAAGAAGACTACTGGAACTGGATGACCTCAGAATGAACTGGATTGTGGTGTTC
ACAAGAAAATCTTAGTTTGTGATGATTACATTGCTTTTGTGTGCCAGTAGTTAGTTTGTGTACAT

In a search of sequence databases, it was found, for example, that a NOV1 nucleic acid sequence, which maps to human chromosome 10, has 989 of 990 bases (99%) identical to a gb:GENBANK-ID:HSM801335|acc:AL133064.1 mRNA from Homo sapiens (Homo sapiens mRNA; cDNA DKFZp434A2417 (from clone DKFZp434A2417; partial cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

In all BLAST alignments herein, the "E-value" or "Expect" value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject ("Sbjct") retrieved from the NOV1 BLAST analysis, *e.g.*, human AL133064.1 mRNA, matched the Query NOV1 sequence purely by chance is $1.1e^{-217}$. The Expect value (E) is a parameter that describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.

The Expect value is used as a convenient way to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, *e.g.*, <http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/>. Occasionally, a string of X's or N's will result from a BLAST search. This is a result of automatic filtering of the query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (*e.g.*, "NNNNNNNNNNNNNN") or the letter "X" in protein sequences (*e.g.*, "XXXXXXXXXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment. (Wootton and Federhen, *Methods Enzymol* 266:554-571, 1996).

The disclosed NOV1 polypeptide (SEQ ID NO:2) encoded by SEQ ID NO:1 has 523 amino acid residues and is presented in Table 1B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1 has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.7300. In other embodiments, NOV1

may also be localized to the endoplasmic reticulum (membrane) with a certainty of 0.6400, the nucleus with a certainty of 0.1800, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site is after position 20 of SEQ ID NO: 2.

Exon linking data for NOV1 can be found below in Example 1. Quantitative gene expression (TaqMan) data for NOV1 can be found below in Example 2. SNP data for NOV1 can be found below in Example 3.

Table 1B. Encoded NOV1 protein sequence (SEQ ID NO:2).

MFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEAAVKVIQSLPKNGSFQPTNEMMLK
FYSFYKQATEGPCKLSRPGFWDPIGRYKWDWSSLDGMTKEEAMIAYVEEMKKI IETMPMTEKVEELLRV
IGPFYEIIVEDKKSGRSSDITSDLGNVLTSAAPNAKTVNGKAESSDSGAESSEEEEAQEEVKGAEQSDNDKKM
MKKSADHKNELEVIVTNGYDKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDH
VEDVTGIQHLTSDSDSEVYCDSEMEQFGQEEESLDSFTSNNGPFQYYLGGHSSQPMENSGFREDIQVPPGNG
NIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRRGRGHRIQHLSEGTGKRQVGSNGD
GERWGSDRGSRGSLNEQIALVLMRLQEDMQNVLQRLQKLETLTALQAKSSTSTLQTAPQPTSQRPSWWPF
EMSPGVLTFAIIWPFIAQWLVLVYQRRRRKLN

The full amino acid sequence of the disclosed NOV1 protein was found to have 443 of 533 amino acid residues (83%) identical to, and 473 of 533 amino acid residues (88%) similar to, the 533 amino acid residue ptrn:SWISSPROT-ACC:P07106 protein from Bos taurus (ENDOZEPINE-RELATED PROTEIN PRECURSOR (MEMBRANE-ASSOCIATED DIAZEPAM BINDING INHIBITOR) (MA-DBI). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV1 is expressed in at least the following tissues: : Brain, Colon, Foreskin, Kidney, Larynx, Lung, Mammary gland/Breast, Ovary, Pancreas, Placenta, Retina, Small Intestine, Spleen, Testis, Thalamus, and Uterus.

The amino acid sequence of NOV1 had high homology to other proteins as shown in Table 1C.

Table 1C. BLASTX results for NOV1

	High Score	Smallest Sum Prob P (N)
Sequences producing High-scoring Segment Pairs:		
patp:AAB48379 Human SEC12 protein sequence (clone ID 2093...	2733	2.4e-284
patp:AAU00399 Human secreted protein, POLY11 - Homo sapie...	2733	2.4e-284
patp:AAB48375 Human SEC8 protein sequence (clone ID 20936...	2727	1.0e-283
patp:AAB81816 Human endozepine-like ENDO6 SEQ ID NO: 23 -...	2687	1.8e-279
patp:AAB81816 Human endozepine-like ENDO6 SEQ ID NO: 23 -...	2687	1.8e-279

The disclosed NOV1 polypeptide also has homology to the amino acid sequences shown in the BLASTP data listed in Table 1D.

Table 1D. BLAST results for NOV1

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect t
CAC24877	Sequence 23 from Patent WO0078802/human	534	518/534 (97%)	520/534 (97%)	3.6e- 284
CAC24873	Sequence 15 from Patent WO0078802/human	536	517/531 (97%)	518/531 (97%)	1.5e- 283
P07106	Endozepine- related protein precursor/bovine	533	443/533 (83%)	473/533 (88%)	9.6e- 243
Q9CW41	1300014E15RIK protein/mouse	504	389/517 (75%)	433/517 (83%)	5.8e- 197
Q9UFB5	Hypothetical 31.5kDa Protein/human	283	282/283 (99%)	283/283 (100%)	3.4e- 153

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 1E. In the ClustalW alignment of the NOV1 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 1E. ClustalW Analysis of NOV1

Information for the ClustalW proteins:

1. NOV1 (SEQ ID NO: 2)
2. CAC24877 (SEQ ID NO: 3)
3. CAC24873 (SEQ ID NO: 4)
4. P07106 (SEQ ID NO: 5)
5. Q9CW41 (SEQ ID NO: 6)
6. Q9UFB5 (SEQ ID NO: 7)

```

NOV1      --MFQFHAGSWESWCCCC--LIPADRPWDRGQHWOLEMADTRSVHETRFEA 47
CAC24877  --MYXFHAGSWESWCCCC--LIPADRPWDRGQHWOLEMADTRSVHETRFEA 47
CAC24873  MLFLSFHAGSWESWCCCC--LIPADRPWDRGQHWOLEMADTRSVHETRFEA 49
P07106    --MFQFHAGSWESWCCCC--LIPADRPWDRGQHWOLEMADTRSVHETRFEA 48
Q9CW41    -----FHAGSWESWCCCC--LIPADRPWDRGQHWOLEMADTRSVHETRFEA 45
Q9UFB5    -----1

NOV1      AVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRY 97
CAC24877  AVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRY 97
CAC24873  AVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRY 99
P07106    AVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRY 98
Q9CW41    AVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRY 95
Q9UFB5    -----1

NOV1      KWDANSSLGDMTKKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYEI 147
CAC24877  KWDANSSLGDMTKKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYEI 147
CAC24873  KWDANSSLGDMTKKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYEI 149
P07106    KWDANSSLGDMTKKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYEI 148
Q9CW41    KWDANSSLGDMTKKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYEI 145
Q9UFB5    -----1

```

5	NOV1	VEDKKSGRSSDTS-----DLGNVLTSPNAKTVNGKAESSDSG	186
	CAC24877	VEDKKSGRSSDTSVRLEKISKLEDLGNVLTSPNAKTVNGKAESSDSG	197
	CAC24873	VEDKKSGRSSDTSVRLEKISKLEDLGNVLTSPNAKTVNGKAESSDSG	199
	P07106	VEDKKSGRSSDTSVRLEKISKLEDLGNVLTSPNAKTVNGKAESSDSG	198
	Q9CW41	VEDKKSGRSSDTS-----DLGNVLTSSNAKTVNGKAESSDSG	183
	Q9UFB5	-----	1
10	NOV1	AESEEEAQQEYKGAEQSDNDKKMMKKSADHKMLEIVTVNGYDKDGFVQD	236
	CAC24877	AESEEEAQQEYKGAEQSDNDKKMMKKSADHKMLEIVTVNGYDKDGFVQD	247
	CAC24873	AESEEEAQQEYKGAEQSDNDKKMMKKSADHKMLEIVTVNGYDKDGFVQD	249
	P07106	AESEEEAQQEYKGAEQSDNDKKMMKKSADHKMLEIVTVNGYDKDGFVQD	248
	Q9CW41	AESEEEAQQEYKGAEQSDNDKKMMKKSADHKMLEIVTVNGYDKDGFVQD	232
	Q9UFB5	-----	1
15	NOV1	IQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTG	286
	CAC24877	IQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTG	297
	CAC24873	IQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTG	299
	P07106	IQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTG	298
	Q9CW41	IQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTG	282
20	Q9UFB5	-----	46
25	NOV1	IQHLTSDSDSEVYCDSEMQFGQESLDSFTSNNGPFQYYLGGHSSQPMEN	336
	CAC24877	IQHLTSDSDSEVYCDSEMQFGQESLDSFTSNNGPFQYYLGGHSSQPMEN	347
	CAC24873	IQHLTSDSDSEVYCDSEMQFGQESLDSFTSNNGPFQYYLGGHSSQPMEN	349
	P07106	IQHLTSDSDSEVYCDSEMQFGQESLDSFTSNNGPFQYYLGGHSSQPMEN	348
	Q9CW41	IQHLTSDSDSEVYCDSEMQFGQESLDSFTSNNGPFQYYLGGHSSQPMEN	319
	Q9UFB5	-----	96
30	NOV1	SGFRFDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNMSGAPHREKRG	386
	CAC24877	SGFRFDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNMSGAPHREKRG	397
	CAC24873	SGFRFDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNMSGAPHREKRG	399
	P07106	SGFRFDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNMSGAPHREKRG	398
	Q9CW41	SGFRFDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNMSGAPHREKRG	369
35	Q9UFB5	SGFRFDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNMSGAPHREKRG	146
40	NOV1	GETDEFSNVRGRG-CHRMQHLSEGTGKRGQVSGGGERWGS DRGSRGSLN	435
	CAC24877	GETDEFSNVRGRG-CHRMQHLSEGTGKRGQVSGGGERWGS DRGSRGSLN	446
	CAC24873	GETDEFSNVRGRG-CHRMQHLSEGTGKRGQVSGGGERWGS DRGSRGSLN	448
	P07106	GETDEFSNVRGRG-CHRMQHLSEGTGKRGQVSGGGERWGS DRGSRGSLN	447
	Q9CW41	GETDEFSNVRGRG-CHRMQHLSEGTGKRGQVSGGGERWGS DRGSRGSLN	419
	Q9UFB5	GETDEFSNVRGRG-CHRMQHLSEGTGKRGQVSGGGERWGS DRGSRGSLN	195
45	NOV1	BQIALVLMRLQEDMONVLRQLKLETLTALQAKSSTSTLTQAPQPTSQRP	485
	CAC24877	BQIALVLMRLQEDMONVLRQLKLETLTALQAKSSTSTLTQAPQPTSQRP	496
	CAC24873	BQIALVLMRLQEDMONVLRQLKLETLTALQAKSSTSTLTQAPQPTSQRP	498
	P07106	BQIALVLMRLQEDMONVLRQLKLETLTALQAKSSTSTLTQAPQPTSQRP	495
	Q9CW41	BQIALVLMRLQEDMONVLRQLKLETLTALQAKSSTSTLTQAPQPTSQRP	466
	Q9UFB5	BQIALVLMRLQEDMONVLRQLKLETLTALQAKSSTSTLTQAPQPTSQRP	245
50	NOV1	SNWPFEMSPGVLTFAIIWPFIAQWLVLVYQRRRRKLN	523
	CAC24877	SNWPFEMSPGVLTFAIIWPFIAQWLVLVYQRRRRKLN	534
	CAC24873	SNWPFEMSPGVLTFAIIWPFIAQWLVLVYQRRRRKLN	536
	P07106	SNWPFEMSPGVLTFAIIWPFIAQWLVLVYQRRRRKLN	533
	Q9CW41	SNWPFEMSPGVLTFAIIWPFIAQWLVLVYQRRRRKLN	504
55	Q9UFB5	SNWPFEMSPGVLTFAIIWPFIAQWLVLVYQRRRRKLN	283

The presence of identifiable domains in NOV1 was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (<http://www.ebi.ac.uk/interpro>). DOMAIN results for NOV1 as disclosed in Tables 1F, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Table 1F and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading or by the sign (!) and “strong” semi-conserved residues are indicated by grey shading or by the sign (+). The “strong” group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Table 1F lists the domain description from DOMAIN analysis results against NOV1. This indicates that the NOV1 sequence has properties similar to those of other proteins known to contain this domain.

Table 1F. Domain Analysis of NOV1				
pfam00887, ACBP, Acyl CoA binding protein.				
NOV1:	41	HETRFEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRYKWD	100	
ACBP:	1	LQEQFEAAAEKVKKLKKN---PSNDELLQLYSLYKQATVGCNTEKPGMFDLKGRAKWD	56	
NOV1:	101	AWSSLGDMTKEEAMIAVVEEMKKIIETMP	129	(SEQ ID NO:2)
ACBP:	57	AWNELKGMSKEEAMKAYIAKVEELIAKYA	85	(SEQ ID NO: 8)

5 NOV2

NOV2 includes a novel endozepine-related precursor-like protein and 17 variants. The disclosed sequences have been named NOV2a-r. A multiple sequence alignment highlighting the differing amino acids is shown below in Table 30. The methods in cloning these variants are described in detail in Example 4.

NOV2a

NOV2a includes a novel endozepine-related protein disclosed below. A disclosed NOV2a nucleic acid of 1747 nucleotides (also referred to as CG51523-05) encoding a novel endozepine-related protein is shown in Table 12A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 36-38. A putative untranslated region upstream from the initiation codon is underlined in Table 12A. The start codon is in bold letters.

Table 12A. NOV2a nucleotide sequence (SEQ ID NO: 21).
<p> ATGTACACAACTAACTACTGGACAACAAAAGCAATGTAATCATCACAACTAAGATT TTCTTGTAACACCACAATCCAGTTCATTCTGAGGTCATCCAGTTCAGTAGTCTTCTTG AGGAAAACACCATTTTCTCCTCAGTTCAGTTTTCTTCTCCTTCTTGATAGTATAAATACAC CAACCACTGTGCAATAAAAGGCCATATGATGGCAAACGTTAGCACACCAGGAGACATCTC GAAGGGCCACCAAGATGGTCTCTGTGAGGTGGGCTGAGGAGCAGTCTGCAATGTTGATGT TGATGATTTTGCCTGCAAAGCAGTCAGCGTTTCCAGTTTCTGCAGTCTCTGAAGGACATT CTGCATGTCTCTGCACTCTCATCAGCACGAGGGCGATCTGCTCATTGAGGCTGCCTCG GGACCCTCTGTGCGAGCCCCAGCGCTCCCCATCACCTCCACTCCCACCTGCCGGCCCTT GGTTCCTTCTCGCTCAAGTGTGTATCCTATGTCCTCTTCTCTTAACATTAGAGAATTC GTCAGTTTCTCCGCCTCGCTTCTCCCGGTGTGGTGTCTCCGCTGTTATTCTGCCATCTTC TCCTCCATGCTTGACTTCACCTTTTCTTCAACTGCAACCACCTGCATATTCCCAATGTT GCCATTTCCAGGAGGTACTTGAATATCTTCACGAAATCCAGAATTTCCATGGGTTGACT GGAATGACCACCCAAGTAATACTGAAATGGTCCATTGTTGGACGTAAAGCTGTCTAAAGA CTCTTCTGTGCTCAAATTGTTCCATAGAATCACAGTAACTTCACGTCTGAAATCGCTTGT CAAATGCTGAATTCTGTAAACATCTTCAACATGATCATCATTTATATCTTGGTGAATGCA AACAGCAGATTTTCCAGTTTGCCCCAAGTTTCATCAATGGGCTTTACTTCTTCAGTGCT TCTGCCATTTCAGGGAAGAACTGGCATGAATGTCATTCTGTATATCTTGAACAAAGCCATC TTTATCATAGCCATTAGTGACAATGACTTCCAAATCTTATGGTCTGCTGACTTCTTCAT </p>


```

CATTCTTATCATTATCACTTTGTTCTGCTCCTTTCACTTCTTCTGGGCCTCTTCTTC
CTCAGACTCGGCTCCACTGTCACTGCTTTAGCTTTACCATTAAACGGTTTTGGCGTTCGG
AGCAGAAGTGAGAACATTACCAAGATCTGAGGTTATATCAGAACTCCTGCCACTCTTTTT
GTCCTCGACAATTTCAAAAATGGACCTATGACACGCAGCAATTCTTCAACTTTCTCAGT
CATTGGCATAGTTTCAATAATCTTTTTTCATTTCTTCAACATATGCAATCATGGCTTCCTC
TTTGGTCATATCACCCAGTGAAGTCCAAGCATCCCATTTATATCTTCCAATAGGATCCCA
AAATCCAGGCCTTGAAAGTTTACAGGTCCTTCAGTTGCCTGCTTATAGAAGCTATAAAA
TTTAAGCATCATTTTCAATTTGTTGGCTGGAATGAACATTCTTCGGCAAACCTGGATCAC
CTTCACGGCCGCTCAAACCTAGTCTCGTGCACGGATCTCGTGTCCGCCATCTCCAGCTG
CCAGTGTGGCCCCGGTCCCAAGGTCTGTGCGCGGAATCAGGCAGCAGCAGCACCAGCT
TTCCCAAGAGCCTGCATGAACTGGAACATGGAGCGCAGCCGCGGATCAACATGCCCCAA
AAGGAGA (SEQ ID NO: 21)

```

In all BLAST alignments herein, the “E-value” or “Expect” value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject (“Sbjct”) retrieved from the NOV1 BLAST analysis, e.g., human CAC24877 mRNA, matched the Query NOV1 sequence purely by chance is $3.7e-284$. The Expect value (E) is a parameter that describes the number of hits one can “expect” to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.

The Expect value is used as a convenient way to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, e.g., <http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/>. Occasionally, a string of X’s or N’s will result from a BLAST search. This is a result of automatic filtering of the query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter “N” in nucleotide sequence (e.g., “NNNNNNNNNNNNNN”) or the letter “X” in protein sequences (e.g., “XXXXXXXXXX”). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment. (Wootton and Federhen, *Methods Enzymol* 266:554-571, 1996).

The disclosed NOV2a polypeptide (SEQ ID NO: 22) encoded by SEQ ID NO: 21 has 523 amino acid residues and is presented in Table 12B using the one-letter amino acid code.

Table 12B. Encoded NOV2a protein sequence (SEQ ID NO: 22).

```
MFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEAAVKVIQSLPKNGS
FQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAVVEE
MKKIIETMPMTEKVEELLRVIGPFYEIVEDKKSGRSSDITSDLGNVLTSAAPNAKTVNGKA
ESSDSGAESEEEEAQEEVKGAEQSDNDKMMKKSADHKNLEVIIVTNGYDKDGFVQDIQND
IHASSSLNGRSTEEVKPIDENLGQTGKSACVCIHQDINDDHVEDVTGIQHLTSDSDSEVYC
DSMEQFGQEEESLDSFTSNNGPFQYYLGGHSSQPMENSGFREDIQVPPGNGNIGNMQVVAV
EGKGEVKHGGEDGRNNSGAPHREKRGGGETDEFSNVRRGRGRHRIQHLSEGTKGRQVGSGGD
GERWGSDRGSRGSLNEQIALVLMRLQEDMQNVLQRLQKLETLTALQAKSSTSTLTAPQP
TSQRPSWWPFEMSPGVLTFAIIWPFIAQWLVLVLYYQRRRRKLN (SEQ ID NO: 22)
```

The full amino acid sequence of the disclosed NOV2a protein was found to have 518 of 534 amino acid residues (97%) identical to, and 520 of 534 amino acid residues (97%) similar to, the 534 amino acid residue ptrn:REMTRMBL-ACC:CAC24877 protein from sequence 23 from patent WO0078802. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV2a is expressed in at least the following tissues: Brain, Colon, Foreskin, Kidney, Larynx, Lung, Mammary gland/Breast, Ovary, Pancreas, Placenta, Retina, Small Intestine, Spleen, Testis, Thalamus, and Uterus.

The amino acid sequence of NOV2a had high homology to other proteins as shown in Table 12C.

Table 12C. BLASTX results for NOV2a

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Prob P(N)
patp:AAM78692 Human protein SEQ ID NO 1354 - Homo sapiens...	2740	5.3e-285
patp:AAB48379 Human SEC12 protein sequence (clone ID 2093...	2733	2.9e-284
patp:AAU00399 Human secreted protein, POLY11 - Homo sapie...	2733	2.9e-284
patp:AAB48375 Human SEC8 protein sequence (clone ID 20936...	2727	1.3e-283
patp:AAB81816 Human endozepine-like ENDO6 SEQ ID NO: 23 -...	2687	2.2e-279

The disclosed NOV2a polypeptide also has homology to the amino acid sequences shown in the BLASTP data listed in Table 12D.

Table 12D. BLAST results for NOV2a

Gene Index/Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
CAC24877	Sequence 23 from Patent WO0078802/human	534	518/534 (97%)	520/534 (97%)	3.7e-284
CAC24873	Sequence 15 from Patent WO0078802/human	536	517/531 (97%)	518/531 (97%)	1.6e-283
P07106	Endozepine-related protein precursor/bovine	533	443/533 (83%)	473/533 (88%)	1.0e-242
Q9CW41	1300014E15RIK Protein	504	389/517 (75%)	433/517 (83%)	6.0e-197
Q9UFB5	Hypothetical 31.5kDa Protein/human	283	282/283 (99%)	283/283 (100%)	3.5e-153

The presence of identifiable domains in NOV2a was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (<http://www.ebi.ac.uk/interpro>). DOMAIN results for NOV2a and its variants as disclosed in Table 30, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Table 30 and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading or by the sign (!) and “strong” semi-conserved residues are indicated by grey shading or by the sign (+). The “strong” group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Table 12E lists the domain description from DOMAIN analysis results against NOV2a. This indicates that the NOV2a sequence has properties similar to those of other proteins known to contain this domain.

Table 12E. Domain Analysis of NOV2a

ACBP (InterPro) Acyl CoA binding protein		
ACBP: domain 1 of 1, from 41 to 129: score 199.7, E = 4.4e-56		
		*->lqedFeaAaekvkkLkknGpvkPSneekLkLYsLYKQATvGDvnter
		++++FeaA+++++L+knG+++P+ne++Lk+Ys+YKQAT+G++++r
NOV2a	41	HETRFEA AVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSR 87
PGmfdlkgrAKWDAWnelkGmSkeeAmkaYIakVeeLiakya<-*		
		PG++D++gr+KWDAW++l++m+keeAm+aY++++++i++++
NOV2a	88	PGFWDPIGRYKWDASSLGDMTKEEAMIAVVEEMKKIETMP 129

NOV2b

In an alternative embodiment, a NOV2 variant is NOV2b of 1432 nucleotides (also referred to as CG51523-05_164786042), shown in Table 13A. A NOV2b variant differs from NOV2a at positions 170, 374, 403, and 493.

Table 13A. NOV2b nucleotide sequence (SEQ ID NO: 23).

AAGCTTGACAGACCTTGGGACCGGGGCCAACACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTT
TGAGGCGGCCGTGAAGGTGATCCAGAGTTTGCCGAAGAATGGTTCATTCCAGCCAACAAATGAAATGATGCTTAAATTTT
ATAGCTTCTATAAGCAGGCAACTGAAGGACCCCTGTAAACTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAA
TGGGATGCTTGGAGTTCACTGGGTGATATGACCAAAGAGGAAGCCATGATTGCATATGTTGAAGAAATGAAAAAGATTAT
TGAAATATGCCAATGACTGAGAAAGTTGAAGAATTGCTGCGTGTCTAGGTCCATTTTATGAAATGTCTGAGGACAAAA
AGAGTGGCAGGAGTTCTGATATAACCTCAGATCTTGGTAATGTTCTCACTTCTACTCCGAACGCCAAAACCGTTAATGGT
AAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAAGGAGCAGAACAAAGTGA
TAATGATAAGAAAATGATGAAGAAGTCAGCAGACCATAAGAATTTGGAAGTCATTGTCTACTAATGGCTATGATAAAGATG
GCTTTGTTTCAAGATATACAGAATGACATTCATGCCAGTTCCTCCCTGAATGGCAGAAGCACTGAAGAAGTAAAGCCCATT
GATGAAAACCTTGGGGCAAACCTGGAAAATCTGCTGTTTGCATTACCAAGATATAAATGATGATCATGTTGAAGATGTTAC
AGGAATTCAGCATTTGACAAGCGATTGAGCAGTGAAGTTTACTGTGATTCTATGGAACAATTTGGACAAGAAGAGTCTT
TAGACAGCTTTACGTCCAACAATGGACCATTTCAGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAATTCCTGGA
TTTCGTGAAGATATTCAAGTACCTCCTGGAAATGGCAACATTGGGAATATGCAGGTGGTTCAGTTGAAGGAAAAGGTGA
AGTCAAGCATGGAGGAGAAGATGGCGGGAATAACAGCGGAGCACCACACCGGGAGAAGCGAGGCGGAGAACTGACGAAT
TCTCTAATGTTAGAAGAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGGCCGGCAGGTGGGAAGTGA
GGTGTATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCAGGAGCAGCTCAACGAGCAGATCGCCCTCGTGCTGATGAGACT
GCAGGAGGACATGCAGAATGTCTTCAGAGACTGCAGAACTGGAAACGCTGACTGCTTGCAGGCAAAATCATCAACAT
CAACATTGCAGACTGCTCCTCAGCCACCTCACAGAGACCATCTTGGTGGCCCTTCGAGATGCCCTCTCGAG (SEQ ID
NO: 23)

Table 13B. Encoded NOV2b protein sequence (SEQ ID NO: 24).

KLDRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRYKWD
AWSSLGDMTKEEAMIAAYVEEMKKI IETMPMTEKVEELLRVIGPFYFIEVDKKSGRSSDITSDLGNVLTSTPNAKTVNGKAES
SDSGAESEEEEEQBEVKGAEQSDNDKMMKKSADHKNLEVI VINGYDKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLGQ
TGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDSEMQFGQESLDSFTSNNGPFQYLLGGHSSQPMENSGFREDIQVP
PGNGNIGNMQVVAVEGKGEVKHGGEDGGNNSGAPHREKRGGTDEFNSVRRGRGHRMQHLSEGTKGRQVGSQGDGERWSDR
GSRGSLNEQIALVLMRLQEDMQNLQRLQKLETLTALQAKSSTSLQTAPQPTSQRPSWWPFEMPSR (SEQ ID NO:
24)

NOV2c

In an alternative embodiment, a NOV2 variant is NOV2c of 1401 nucleotides (also referred to as CG51523-05_164732479), shown in Table 14A. A NOV2c variant differs from NOV2a at positions 71, 170, 313, and 403, and by an insertion of 11 amino acids at positions 161-162.

Table 14A. NOV2c nucleotide sequence (SEQ ID NO: 25).

AAGCTTACTAGGTTTGGAGCGGCCGTGAAGGTGATCCAGAGTTTGCCGAAGAATGGTTCATTCCAGCCAACAAATGAAAT
GATGCTTAAATTTTATAGCTTCTATAAGCAGGCAACTGAAGGACCCCTGTAAACTTTCAAGGCCTGGATTTTGGGATCCTA
TTGGAAGATATAAATGGGATGCTTGGAGTTCACTGGGTGATATGACCAAAGGGGAAGCCATGATTGCATATGTTGAAGAA
ATGAAAAAGATTATTGAAACTATGCCAATGACTGAGAAAGTTGAAGAATTGCTGCGTGTCTAGGTCCATTTTATGAAAT

TGTCGAGGACAAAAAGAGTGGCAGGAGTTCTGATATAACCTCAGTCCGACTGGAGAAAATCTCTAAATGTTTAGAAGATC
 TTGGTAATGTTTCTCACTTCTACTCCGAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCT
 GAGGAAGAAGAGGCCCAAGAAGAAGTGAAGGAGCAGAACAAAGTGATAATGATAAGAAAATGATGAAGAAGTCAGCAGA
 CCATAAGAATTTGGAAGTCATTGTCACTAATGGCTATGATAAAGATGGCTTTGTTTCAGGATATACAGAATGACATTCATG
 CCAGTTCTTCCCTGAATGGCAGAAGCACTGAAGAAGTAAAGCCCATTTGATGAAAACCTGGGGCAAACCTGGAAAATCTGCT
 GTTTCGATTACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTGAGCATTGACAAGCGATTGAGACAG
 TGAAGTTTACTGTGATTCTATGGAACAATTGGACAAGAAGAGTCTTTAGACAGCTTTACGTCCAACAATGGACCATTTC
 AGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAATCTGGATTTCGTGAATATATTCAAGTACCTCCTGGAAAT
 GGCAACATTGGGAATATGCAGGTGGTTGCAGTTGAAGGAAAAGGTGAAGTCAAGCATGGAGGAGAAGATGGCAGGAATAA
 CAGCGGAGCACCACACCGGGGAGAAGCGAGGCGGAGAACTGACGAATTCTCTAATGTTAGAAGAGGAAGAGGACATAGGA
 TGCAACACTTGAGCGAAGGAACCAAGGGCCGGCAGGTGGGAAGTGGAGGTGATGGGGAGCGCTGGGGCTCCGACAGAGGG
 TCCCGAGGCAGCTCAATGAGCAGATCGCCCTCGTGTGATGAGACTGCAGGAGGACATGCAGAATGTCCTTCAGAGACT
 GCAGAACTGGAAACGCTGACTGCTTTGCAGGCAAAATCATCAACATCAACATTGCAGACTGCTCCTCAGCCCACCTCAC
 AGAGACCATCTTGGTGGCCCTTCGAGATGTCTCCTCTCGAG (SEQ ID NO: 25)

Table 14B. Encoded NOV2c protein sequence (SEQ ID NO: 26).

KLTRFEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRYKWDWSSSLGDMTKGEAMIAYVEE
 MKKIIETMPMTEKVEELLRVIGPFYEIVEDKKSGRSSDITSVRLEKISKLEDLGNVLTSTPNAKTVNGKAESSDSGAE
 EEEEEQEEVKGAEQSDNDKMMKKSADHKNLEIVITNGYDKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLGQTGKSA
 VCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDMEQFQGEESLDSFTSNNGPFQYLLGGHSSQPMENSGFREYIQVPPGN
 NIGNMQVVAVEGKGEVKGHGGEDGRNNSGAPHREKRGGETDEFNSNRRGRHRMQHLSEGTKGRQVGSAGGDERWGS
 RGSLSNEQIALVLMRLQEDMQNVLRQLQKLETLTALQAKSSTLTQTAPQPTSQRPSWPFEMSPLE (SEQ ID NO: 26)

NOV2d

In an alternative embodiment, a NOV2 variant is NOV2d of 1401 nucleotides (also
 5 referred to as CG51523-05_164732506), shown in Table 15A. A NOV2d variant differs from
 NOV2a at positions 170, 292, and 403, and by the insertion of 11 amino acids at position 161-
 162.

Table 15A. NOV2d nucleotide sequence (SEQ ID NO: 27).

AAGCTTACTAGGTTTGGAGCGCCGTGAAGGTGATCCAGAGTTTGCCGAAGAATGGTTCATTCCAGCCAACAAATGAAAT
 GATGCTTAAATTTTATAGCTTCTATAAGCAGGCAACTGAAGGACCTGTAAACTTTCAAGGCCTGGATTTTGGGATCCTA
 TTGGAAGATATAAATGGGATGCTTGGAGTTCACTGGGTGATATGACCAAAGAGGAAGCCATGATTGCATATGTTGAAGAA
 ATGAAAAAGATTATGAAACTATGCCAATGACTGAGAAAGTTGAAGAATTGCTGCGTGTATAGGTCCATTTTATGAAAT
 TGTCGAGGACAAAAAGAGTGGCAGGAGTTCTGATATAACCTCAGTCCGACTGGAGAAAATCTCTAAATGTTTAGAAGATC
 TTGGTAATGTTTCTCACTTCTACTCCAAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCT
 GAGGAAGAAGAGGCCCAAGAAGAAGTGAAGGAGCAGAACAAGTGATAATGATAAGAAAATGATGAAGAAGTCAGCAGA
 CCATAAGAATTTGGAAGTCATTGTCACTAATGGCTATGATAAAGATGGCTTTGTTTCAGGATATACAGAATGACATTCATG
 CCAGTTCTTCCCTGAATGGCAGAAGCACTGAAGAAGTAAAGCCTATTGATGAAAACCTGGGGCAAACCTGGAAAATCTGCT
 GTTTCGATTACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTGAGCATTGACAAGCGATTGAGACAG
 TGAAGTTTACTGTGATTCTATGGAACAATTTGGACAAGAAGAGTCTTTAGACAGCTTTACGTCCAACAATGGACAATTTT
 AGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAATCTGGATTTCGTGAAGATATTCAAGTACCTCCTGGAAAT
 GGCAACATTGGGAATATGCAGGTGGTTGCAGTTGAAGGAAAAGGTGAAGTCAAGCATGGAGGAGAAGATGGCAGGAATAA
 CAGCGGAGCGCCACACCGGGAGAAGCGAGGCGGAGAACTGATGAATTCTCTAATGTTAGAAGAGGAAGAGGACATAGGA
 TGCAACACTTGAGCGAAGGAACCAAGGGCCGGCAGGTGGGAAGTGGAGGTGATGGGGAGCGCTGGGGCTCCGACAGAGGG
 TCCCGAGGCAGCTCAATGAGCAGATCGCCCTCGTGTGATGAGACTGCAGGAGGACATGCAGAATGTCCTTCAGAGACT
 GCAGAACTGGAAAACGCTGACTGCTTTGCAGGCAAAATCATCAACATCAACATTGCAGACTGCTCCTCAGCCCACCTCAC
 AGAGACCATCTTGGTGGCCCTTCGAGATGTCTCCTCTCGAG (SEQ ID NO: 27)

KLTRFEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRYKWDAWSSLDGMTKEEAMIAYVEEM
KKIIETMPMTEKVEELLRVIGPFYEIVEDKKSGRSSDITSVRLEKISKCLEDLGNVLTSTPNAKTVNGKAESSDSGAESEE
EAQEEVKGAEQSDNDKKMMKKSADHKNLEVIITNGYDKDGFVQDIQNDIHASSLSNGRSTEEVKPIDENLGQTGKSAVCIH
QDINDDHVEDVTGIQHLLTSDSDSEVYCDSEMQFGQEESLDSFTSNNGQFYLYLGGHSSQPMGNSFREDIQVPPGNINIGNM
QVVAVEGKGVEKHGDEGRNNSGAPHREKRGGETDEFSNVRRGGRHMQHLSEGTKRGVQSGGDERWGSRGSRLNE
QIALVLMRLQDPMNVLORLOKLETLTALQAKSSSTLTQTAPQPTSQRPSSWWPFEMSPLE (SEQ ID NO: 28)

In an alternative embodiment, a NOV2 variant is NOV2e of 1401 nucleotides (also referred to as CG51523-05_164732693), shown in Table 16A. A NOV2e variant differs from NOV2a at the protein level at positions 170 and 403, and by the insertion of 11 amino acids at position 161-162.

AAGCTTACTAGGTTTGAGGCGGCCGTGAAGGTGATCCAGAGTTTGCCGAAGAATGGTTCATCCAGCCAACAAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGCAGGCAACTGAAGGACCCGTGTAACCTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGTTCACCTGGGTGATATGACCAAGAGGAGGCCATGATTCGATATGTTTGAAGAAATGAAAAAGATTATTGAAACTATGCAACTGACAGTGAAGAAGTGAAGAATTCGTCGCTGTCATAGGTCCTATTTTATGAAATGTTCGAGGACAAAAGAGTGGCAGGAGTTCGTGATAAACCCTCAGTCCGCACTGGAGAAAATCTCTAAATGTTTGAAGATCTTGGTAATGTTCTCACTTCTACTCCAAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAGTGAAAGGAGCAGAACAAATGATAATGATAAGAAAATGATGAAGAAGTCAGCAGACCATAGAAGAAATTTGGAAGTCATTGTCACTAATGGCTATGATAAAGATGGCTTGTGTTCCAGGATATACAGAATGACATTCATGCCAGTCTCTCCCTGAATGGCAGAAGCACTGAAGAAGTAAAGCCCATGTATGAAGAACTGGGGCAAACCTGGAAGAAATCTGCTGTTTGGCATTCACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTCAGCATTTGACAAGCGATTTCAGACAGTGAAGTTTACTGTGATTCTATGGAACAATTTGGACAAGAAGAGTCTTTAGACAGCTTTACGTCCACAATGGACCATTTACGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAATTTCTGGAAATTCGTGAAGATATTCAAGTACCTCCTGGAAATGGCAACACTTGGGAATATGCAAGGTGGTTGCAGTTGAAGAAAAGGTGAAGTCAAGCATGGAGGAGAAGATGGCAGGAATAACAGCGGAGCACCAACCCGGGGAAGCAGGCGGAGGAGTACGAAATCTCTTAATGTTTGAAGAGGAAGAGGCATATGGAATGCAACACTTGAGCGAAGGAACCAAGGGCCGGCAGGTGGGAAGTGGAGGTGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCAGAGCGAGCTCAATGAGCAGATCGCCCTCGTGCTGATGAGACTGCAAGGAGGACATGCAGAATGTCCTTCAGAGACTGCAGAAAACCTGGAACCGTGACTGCTTTGCGAGCAAAATCATCAACATCAACATGACACTGCTCCTCAGCCCACCTCACAGAGACCATCTTGGTGGCCCTTCGAGATGTCTCTCTCGAG (SEQ ID NO: 29)

KLTRFEAAVKVIQSLPKNGSGFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRYKWDASSLGDMTKEEAMIAVVEEM
KKIIETMPMTEKVEELLRVIGPFYIEVDKKSGRSSDITSVRLEKISKCLEDLAAAAEQEEVKGAEQSDNDKMMKKSADH
KNLEVIVTNGYDKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLQGTGKSAVCIHQDINDHDHVEDVTGIQHLTSDSSEV
YCDMEQFGQEESSLDSFTSNNGPFQYYLGHHSSQPMENSGFREDIQVPPGNNIGNMQVVAVEGKGEVKHGGEDGRNNSGAP
HREKRGGETDEFSNVRNRGRHRMQLHSEGTGCRGVSGSGDGERWSDGRSGRLNBEQIALVLMRLQEDMQNVLQRLQKLET
LTLAOLAKSSTLTLOTAPOPTSORPSWPFEMPLE (SEQ ID NO: 30)

In an alternative embodiment, a NOV2 variant is NOV2f of 1368 nucleotides (also referred to as CG51523-05_164732709), shown in Table 17A. A NOV2f variant differs from NOV2a at the protein level at positions 170, 403, 449, and 485.

Table 17A. NOV2f nucleotide sequence (SEQ ID NO: 31).

AAGCTTACTAGGTTTGAGGCGGCCGTGAAGGTGATCCAGAGTTTGCCGAAGAATGGTTCATTCCAGCCACAAATGAAAT GATGCTTAAATTTTATAGCTTCTATAAGCAGGCAACTGAAGGACCCCTGTAACCTTTCAAGGCCTGGATTTTGGGATCCTA TTGGAAGATATAAATGGGATGCTTGGAGTTCACTGGGTGATATGACCAAAGAGGAAGCCATGATTGCATATGTTGAAGAA ATGAAAAAGATTATTGAACTATGCCAATGACTGAGAAAAGTTGAAGAATTGCTGCGTGTGCATAGGTCATTTTATGAAAT TGTCGAGGACAAAGGAAGTGGCAGGAGTTCTGATATAACCTCAGATCTTGGTAATGTTCTCACTTCTACTCCGAACGCCA AAACCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAAAGGA GCAGAACAAAGTGATAATGATAAGAAAATGATGAAGAAGTCAGCAGACCATAAGAATTTGGAAGTCATTGTCTACTAATGG CTATGATAAAGATGGCTTTGTTTCAGGATATACAGAATGACATTCATGCCAGTTCCTCCCTGAATGGCAGAAGCACTGAAG AAGTAAAGCCCATTTGATGAAAACCTTGGGGCAAACTGGAAAACTCTGCTGTTTGCATTCACCAAGATATAAATGATGATCAT GTTGAAGATGTTACAGGAATTCAGCATTGTGACGAGCGATTTCAGACAGTGAAGTTTACTGTGATTCTATGGAACAATTTGG ACAAGAAGAGTCTTTAGACAGCTTTACGTCCAACAATGGACCATTTAGTATTACTTGGGTGGTTCATTCCAGTCACCCCA TGAAAAATTTCTGGATTTCCTGTAAGATATTCAAGTACCTCCTGGAAATGGCAACATTGGGAATATGCAGGTGGTTGCAGTT GAAGGAAAAGGCGAAGTCAAGCATGGAGGAGAAGATGGCAGGAATAACAGCGGAGCACCACACCGGGAGAAGCGAGGCGG AGAAACTGACGAATTTCTCTAATGTTAGAAGAGGAAGAGGACATAGGATGCAACAATTGAGCGAAGGAACCAAGGCGCCGC AGGTGGGAAGTGGAGGTGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGCCCTC GTGCTGATGAGACTGCAGGAGGACATACAGAATGTCCTTCAGAGACTGCAGAACTGGAAACGCTGACTGCTTTTCGAGGC AAATCATCAACATCAACATTTGCAGACTGCTCTCAGCCCACTCAGAGATCATCTTGGTGCCCTTCGAGATGTCTC CTCTCGAG (SEQ ID NO: 31)

Table 17B. Encoded NOV2f protein sequence (SEQ ID NO: 32).
<p>KLTRFEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQATEGPCKLSRPGFWDPIGRYKWDAWSSLGDMTKEEAMIAIYVEEM KKI IETMPMTEKVEELLRVIGPFYIEIVEDKKSGRSSDITSDLGNVLTSTPNAKTVNGKAESSDSCAEESEEEEAQEEVKGEQ SDNDKMMKKSADHKNLEIVITNGYDKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHEDVT GIQHLTSDSDSEVYCDMSMEQFGQEESELDSFTSNNGPFQYYLGGHSSQPMENSGFREDIQVPPGNGNIGNMQVVAVGKGEVK HGGEDGRNNSGAPHREKRGGETDEFSNVRRGRGHRMQHLSEGTKGRQVGGSGDGERWGSDRGSRGSLNEQIALVLMRLQED IQNVLQRLQKLETLTALQAKSSTSTLQTAPQPTSQRSSWWPFEMSPLE (SEQ ID NO: 32)</p>

In an alternative embodiment, a NOV2 variant is NOV2g of 1586 nucleotides (also referred to as CG51523-05_164718189), shown in Table 18A. A NOV2g variant differs from NOV2a by 2 amino acids at positions 170 and 403.

Table 18A. NOV2g nucleotide sequence (SEQ ID NO: 33).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGCTGCTGCTGCTGATTCCCGCCGACAGACCT
TGGGACCGGGGCCAACACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGAGGCGGCCGTGAA
GGTGATCCAGAGTTTGGCCGAGAATGGTTTCATTCCAGCCCAACAAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC
AGGCAACTGAAGGACCCCTGTAAATCTTCAAGCGCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT
TCACTGGGTGATATGACCAAAGAGGAAAGCAGTCATGTTGCATATGTTGAAGAAATGAAAAGATTATGAACTATGCCAAT
GACTGAGAAAGTTGAAGAATTGCTGCGTGTCATAGGTCCATTTTATGAAATTGTCGAGGACAAAAAGAGTGGCAGGAGTT
CTGATATAACCTCAGATCTTGGTAATGTTTCTCACTTCTACTCCGAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGT
GACAGTGGAGCCGAGTCTGAGGAGAAGAGGGCCCAAGAAGAAGTGAAAGGAGCAGAACAAAGTGATAATGATAAGAAAAAT
GATGAAGAAGTCAAGCAGACGATAAGAATTTGGAAGTCATTGTCACATAATGGCTATGATAAAGATGGCTTTGTTCCAGGATA
TACAGAATGACATTCATGCCAGTTCTTCCCTGTAATGGCAGAGAAGCACTGAAGAAGTAAAGCCCATTTGATGAAACATTTGGGG
CAAACCTGGAAGAAATCTGCTGTTTGCATTACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTCAGCATT
GACAAGCGATTGAGACAGTGAAGTTTACTGTGATTCTATGGAACAATTTGGACAAGAAGAGTCTTTAGACAGCTTTACGT
CCAACAATGACACCAATTCAGTATTACTTGGGTGGTCAATTCAGTCAACCCATGGAAATTCGTGATTTCTGTGAAGATATT
CAAGTACCTCCTGGAAATGGCAACATTGGGAAATATGCAGGTGGTTGCGAGTTGAAGGAAAAGGTGAAGTCAAGCATGGAG
AGAAGATGGCAGGAATAACACGCGGAGCACCACCGGGAGAGCGGAGCGGAGAACTGACAAATCTCTAATGTTAGAA
GAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGGCCGGCAGGTGGGAAGTGGAGGTGATGGGGAGCGC
TGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGCCCTCGTGCTGATGAGACTGCAGGAGGACATGCA
GAATGTCCTTCAGAGACTGCAGAACTGGAACCGTGACTGCTTTGCAGGCCAAAATCATCAACATCAACATTGCAGACTG
CTCCTCAGCCACCTCACAGAGACCATCTTGGTGGCCCTTCGAGATGTCTCCTGGTGTGCTAACGTTTGCCATCATATGG
CCTTTATTGCACAGTGGTTGGTGTATTTATACTATCAAAGAAGGAGAAGAAAACCTGAACCTCGAG (SEQ ID NO:
33)

Table 18B. Encoded NOV2g protein sequence (SEQ ID NO: 34).

ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQ
ATEGPCKLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYIEVEDKKSGRSS
DITSDLGNVLTSTPNAKTVNGKAESSDSGAEESEEEAQEEVKGAEQSDNDKKMMKKSADHKNLEVIIVTNGYDKDGFVQDI
QNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDSEMEQFGQEEESLDSFTS
NNGPFQYYLGGHSSQPMENSGFREDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRR
GRGHRMQHLSEGTKGRQVGSAGDGERWGS DRGSRGSLNEQIALVLMRLQEDMQNVLQRLQKLETLTALQAKSSTSTLQTA
PQPTSQRPSWWPFEMSPGVLTFAIIWPFIAQWLVLVYLYQRRRRKLNLE (SEQ ID NO: 34)

NOV2h

In an alternative embodiment, a NOV2 variant is NOV2h of 1618 nucleotides (also referred to as CG51523-05_164718193), shown in Table 19A. A NOV2h variant differs from NOV2a by the first twenty amino acids, and the 3 amino acids at positions 170, 182 and 403. In addition, NOV2h differs from NOV2a by the insertion of eleven amino acids at position 161-162.

Table 19A. NOV2h nucleotide sequence (SEQ ID NO: 35).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGCTGCTGCTGCCTGATCCGCCGACAGACCTT
GGGACCGGGGCCAACACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGAGGCGGCGGTGAAG
GTGATCCAGAGTTTGCCGAAGAATGGTTCATTCCAGCCAACAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGCA
GGCAACTGAAGGACCTGTAAACTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGTT
CACTGGGTGATATGACCAAAGAGGAAGCCATGATTGCATATGTTGAAGAAATGAAAAGATTATTGAACTATGCCAATG
ACTGAGAAAGTTGAAGAATTGCTGCGTGTCTAGGTCCATTTTATGAAATTGTGAGGACAAAAGAGTGCCAGGAGTTC
TGATATAACCTCAGTCCGACTGGAGAAAATCTCTAAATGTTTGAAGATCTTGGTAATGTTCTCACTTCTACTCCAAACG
CCAAAACCGTTAATGGTAAAGCTGAAGGCAGTGACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAAA
GGAGCAGAAACAAAGTGATAATGATAAGAAAATGATGAAGAAGTCAGCAGACCATAAGAATTGGAAGTCATTGTCACTAA
TGGCTATGATAAAGATGGCTTTGTTTCAGGATATACAGAATGACATTCAATGCCAGTTCTTCCCTGAATGGCAGAGCACTG
AAGAAGTAAAGCCCATTGATGAAAACCTTGGGGCAAACCTGGAATCTGCTGTTTGCATTACCAAGATATAAATGATGAT
CATGTTGAAGATGTTACAGGAATTCAGCATTGACAAGCGATTGACAGAGTGAAGTTTACTGTGATTCTATGGAACAATT
TGGACAAGAAGAGTCTTTAGACAGCTTTACGTCCAACAATGGACCATTTCAGTATTACTTGGGTGGTCATTCCAGTCAAC
CCATGGAATAATTCGATTTCGTGAAGATATTCAAGTACCTCCTGGAATGGCAACATTGGGAATATGCAGGTGGTTGCA
GTTGAAGGAAAAGGTGAAGTCAAGCATGGAGGAGAAGATGGCAGGAATAACAGCGGAGCACCACACCGGGAGAAGCGAGG
CGGAGAACTGACGAATTCTTAATGTTAGAAGAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGGCC
GGCAGGTGGGAAGTGGAGGTGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGCC
CTCGTGCTGATGAGACTGCAGGAGGACATGCAGAATGTCCTTCAGAGACTGCAGAACTGGAAACGCTGACTGCTTTGCA
GGCAAAATCATCAACATCAACATTGCAGACTGCTCCTCAGCCACCTCAGAGAGACCATCTTGGTGGCCCTTCGAGATGT
CTCCTGGTGTGCTAACGTTTGGCATCATATGGCCTTTTATTGCACAGTGGTGGTGTATTATACTATCAAAGAAGGAGA
AGAAAACCTGAACCTCGAG (SEQ ID NO: 35)

Table 19B. Encoded NOV2h protein sequence (SEQ ID NO: 36).

SFHHVPVSCRLGLKLVLLLPDSADRPWDRGQHWQLEMADTRSVHETRFEEAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQA
TEGPCKLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYIEVEDKKSGRSSDI
TSVRLEKISKCLEDLGNVLTSTPNAKTVNGKAEGSDSGAEESEEEAQEEVKGAEQSDNDKKMMKKSADHKNLEVIIVTNGYD
KDGFPQDIQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDSEMEQFGQEE
SLDSFTSNNGPFQYYLGGHSSQPMENSGFREDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDE
FSNVRRGRGHRMQHLSEGTKGRQVGSAGDGERWGS DRGSRGSLNEQIALVLMRLQEDMQNVLQRLQKLETLTALQAKSSTS
TLQTAPQPTSQRPSWWPFEMSPGVLTFAIIWPFIAQWLVLVYLYQRRRRKLNLE (SEQ ID NO: 36)

NOV2i

In an alternative embodiment, a NOV2 variant is NOV2i of 1586 nucleotides (also referred to as CG51523-05_164718197), shown in Table 20A. A NOV2i variant differs from NOV2a by 4 amino acids at positions 170, 403, 422 and 466.

Table 20A. NOV2i nucleotide sequence (SEQ ID NO: 37).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGTCTGCTGCTGCCTGATTCCCGCCGACAGACCT
TGGGACCGGGGCCAACACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGAGGCGGCCGTGAA
GGTGTATCCAGAGTTTGCCGAAGAATGGTTTCATTCAGCCCAACAAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC
AGGCAACTGAAGGACCTGTAACTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT
TCACTGGGTGATATGACCAAGAGGAAGCCATGATTGCATATGTTGAAGAAATGAAAAAGATTATTGAAACTATGCCAAT
GACTGAGAAAGTTGAAGAATTGCTGCGTGTATAGGTCCATTTTATGAAATTGTGAGGACAAAAAGAGTGGCAGGAGTT
CTGATATAACCTCAGATCTTGGTAATGTTCTCACTTCTACTCCGAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGT
GACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAAGGAGCAGAACAAAGTGATAATGATAAGAAAAAT
GATGAAGAAGTCAGCAGACCATTAAGAATTTGGAAGTCATTGTCTACTAATGGCTATGATAAAGATGGCTTTGTTTCAAGATA
TACAGAATGACATTCATGCCAGTTCCTCCCTGAATGGCAGAAGCACTGAAGAAGTAAAGCCATTGATGAAAACTTGGGG
CAAACCTGGAATCTGCTGTTTGCATTACCAAGATATAAATGATGATCATGTTGAAGATGTACAGGAATTGAGCATTT
GACAAGCGATTGAGCAGTGAAGTTTACTGTGATTCTATGGAACAAATTTGGACAAGAAGAGTCTTTAGACAGCTTTACGT
CCAACAATGGACCATTTTCAGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAAATCTGGATTTCGTGAAGATATT
CAAGTACCTCCTGGAAATGGCAACATTGGGAATATGCAGGTGGTTGCAGTTGAAGGAAAAGGTGAAGTCAAGCATGGAGG
AGAAGATGGCAGGAATAACAGCGGAGCACCACACCGGAGAGCGAGGCGGAGAACTGACGAATTCTCTAATGTAGAA
GAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGGCCGCAGGTGGGAAGTGGAGGTGATGGGGGGCGC
TGGGGCTCCGACAGAGGGTCCCCGAGGCAGCCTCAATGAGCAGATCGCCCTCGTGTCTGATGAGACTGCAGGAGGACATGCA
GAATGTCTTCAGAGACTGCAGAACTGGAAACGCTGACTGCTTTGCGGGCAAAATCATCAACATCAACATTGCAGACTG
CTCCTCAGCCCACCTCACAGAGACCATCTTGGTGGCCCTTCGAGATGTCTCTGCTGTGCTAACGTTTGCCATCATATGG
CCTTTTATTGCACAGTGGTTGGTGTATTTATACTATCAAAGAAGGAGAAGAACTGAACCTCGAG (SEQ ID NO:
37)

5

Table 20B. Encoded NOV2i protein sequence (SEQ ID NO: 38).

ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQA
TEGPCKLSRPGFWDPIGRYKDAWSSLGDMTKEEAMIAVVEEMKKIETMPMTEKVEELLRVIGPFYIEVEDKKSGRSSDI
TSDLGNVLTSTPNAKTVNGKAESSDSGAEESEEEAQEEVKGAEQSDNDKKMMKKSADHKNLEVIVTNGYDKDGFVQDIQND
IHASSSLNGRSTEEVKPIDENLQGTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDSEMQFGQESLDSFTSNNGP
FQYYLGGHSSQPMENSGFREDIQVPPNGNIGNMQVAVVEGKGEVKGHGEDGRNNSGAPHREKRGGETDEFNSVRRGRGHR
MQHLSEGTKGRQVSGGGDGRWGSDRGSRGSLNEQIALVLMRLQEDMQNVLRQLQKLETLTALRAKSSTSTLQTPAQPTSQ
RPSWWPFEMSPGVLTFAIIWPFIAQWLVLVLYYQRRRRKLNLE (SEQ ID NO: 38)

NOV2j

In an alternative embodiment, a NOV2 variant is NOV2j of 1517 nucleotides (also referred to as CG51523-05_164718205), shown in Table 21A. A NOV2j variant differs from NOV2a by 4 amino acids at positions 35, 121, 170 and 403, and by a deletion of twenty-three amino acids at position 350.

Table 21A. NOV2j nucleotide sequence (SEQ ID NO: 39).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGTCTGCTGCTGCCTGATTCCCGCCGACAGACCT
TGGGACCGGGGCCAACACTGGCAGCTGGAGATGGTGACACGAGATCCGTGCACGAGACTAGGTTTGAGGCGGCCGTGAA
GGTGTATCCAGAGTTTGCCGAAGAATGGTTTCATTCAGCCCAACAAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC
AGGCAACTGAAGGACCTGTAACTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT
TCACTGGGTGATATGACCAAGAGGAAGCCATGATTGCATATGTTGAAGAAGTGAAGAAAGATTATTGAAACTATGCCAAT
GACTGAGAAAGTTGAAGAATTGCTGCGTGTATAGGTCCATTTTATGAAATTGTGAGGACAAAAAGAGTGGCAGGAGTT
CTGATATAACCTCAGATCTTGGTAATGTTCTCACTTCTACGCCAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGT
GACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAAGGAGCAGAACAAAGTGATAATGATAAGAAAAAT

GATGAAGAAGTCAGCAGACCATAAGAATTTGGAAGTCATTGTCACCTAATGGCTATGATAAAGATGGCTTTGTTTCAGGATA
TACAGAATGACATTCATGCCAGTTCTTCCCTGAATGGCAGAAGCACTGAAGAAGTAAAGCCCATTTGATGAAAACCTTGGGG
CAAACCTGGAAAATCTGCTGTTTGCATTACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTCAGCATTT
GACAAGCGATTGACAGCTGAAGTTTACTGTGATTCTATGGAACAATTTGGACAAGAAGAGTCTTTAGACAGCTTTTACGT
CCAACAATGGACCATTTTCAGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAATCTGGATTTCTGTAAGATATT
CAAGTACCTCTCGAAATGGCAGGAATAACAGCGGAGCACCACACCGGGAGAAGCGAGGCGGAGAACTGACGAATTTCTC
TAATGTTAGAAGAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGGCCGGCAGGTGGGAAGTGGAGGTG
ATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCGAGGCAGCTCAATGAGCAGATCGCCCTCGTGCTGATGAGACTGCAG
GAGGACATGCAGAATGTCCTTCAGAGACTGCAGAACTGGAAACGCTGACTGCTTTGCAGGCAAAATCATCAACATCAAC
ATTGCAGACTGCTCCTCAGCCACCTCACAGAGACCATCTTGGTGGCCCTTCGAGATGTCTCCTGGTGTGCTAACGTTG
CCATCATATGGCCTTTTATTGCACAGTGGTTGGTGTATTTATACTATCAAAGAAGGAGAAGAAAACCTGAACCTCGAG
(SEQ ID NO: 39)

Table 21B. Encoded NOV2j protein sequence (SEQ ID NO: 40).

ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMVDTRSVHETRFEEAVKVIQSLPKNSFQPTNEMMLKFYSFYKQ
ATEGPCKLSRPGFWDPIGRYKWDWSSLDGMTKEEAMIAVVEEVKKI IETMPMTEKVEELLRVIGPFYIEVEDKKSGRSS
DITSDLGNVLTSTPNAKTVNGKAESSDSGAEESEEEAQEEVKGAEQSDNDKMMKKSADHKNLEIVITNGYDKDGFVQDI
QNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDMEQFGQEESSLDSFTS
NNGPFQYLLGGHSSQPMENSGFREDIQVPPNGNRNNSGAPHREKRGGETDEFSNVRRGRHRMQHLSEGTGKRGVSGGGD
GERWGSDRGSRGSLNEQIALVLMRLQEDMQNVLRQLQKLETLTALQAKSSTSTLQTAPQPTSQRPSWWPFEMSPGVLTF
I IWPFI AQWLVLVLYYQRRRRKLNLE (SEQ ID NO: 40)

NOV2k

In an alternative embodiment, a NOV2 variant is NOV2k of 1361 nucleotides (also
5 referred to as CG51523-05_164718209), shown in Table 22A. A NOV2k variant differs from
NOV2a by 68 amino acid deletion at position 208 and 2 amino acid changes. In addition, at
position 162, an 11 amino acid sequence replaces an 18 amino acid sequence.

Table 22A. NOV2k nucleotide sequence (SEQ ID NO: 41).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGCTGCTGCTGCCTGATTCCCGCCGACAGACCT
TGGGACCGGGGCCAACACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGGAGCGCCGTGAA
GGTGATCCAGAGTTTGCCGAAGAATGGTTTCATTCCAGCCAACAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC
AGGCAACTGAAGGACCCTGTAACTTTCAAGGCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT
TCACTGGGTGATATGACCAAAGAGGAAGCCATGATTGCATATGTTGAAGAAATGAAAAGATTATTGAAACTATGCCAAT
GACTGAGAAAGTTGAAGAATTGCTGCGTGTCTAGGTCCATTTTATGAAATTGTCGAGGACAAAAGAGTGGCAGGAGTT
CTGATATAACCTCAGTCCGACTGGAGAAAATCTCTAAATGTTTGAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCTGAG
GAAGAAGAGGCCCCAAGAAGAAGTGAAGGAGCAGAACAAAGTGATAATGATATAAATGATGATCATGTTGAAGATGTTAC
AGGAATTGAGCATTTGACAAGCGATTGACAGAGTGAAGTTTACTGTGATTCTATGGAACAATTTGGACAAGAAGAGTCTT
TAGACAGCTTTACGTCCAACAATGGACCATTTTCAGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAATCTGGA
TTTCGTGAAGATATTCAAGTACCTCCTGGAAATGGCAACATTGGGAATATGCAGGTGGTTGCAGTTGAAGGAAAAGGTGA
AGTCAAGCATGGAGGAGAAGATGGCAGGAATAACAGCGAGCGCCACACCGGGAGAAGCGAGGCGGAGAACTGATGAAT
TCTCTAATGTTAGAAGAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGGCCGGCAGGTGGGAAGTGA
GGTGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGCCCTCGTGCTGATGAGACT
GCAGGAGGACATGCAGAATGTCCTTCAGAGACTGCAGAACTGGAAACGCTGACTGCTTTGCAGGCAAAATCATCAACAT
CAACATTGCAGACTGCTCCTCAGCCACCTCACAGAGACCATCTTGGTGGCCCTTCGAGATGTCTCCTGGTGTGCTAACG
TTTGCCATCATATGGCCTTTTATTGCACAGTGGTTGGCGTATTTATACTATCAAAGAAGGAGAAGAAAACCTGAACCTCGA
G (SEQ ID NO: 41)

Table 22B. Encoded NOV2k protein sequence (SEQ ID NO: 42).

ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMA DTRSVHETRFEEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQ
ATEGPKCLSRPGFWDPIGRYKWDAWSSLGDMTKEEAMIA YVEEMKKI IETMPMTEKVEELLRVIGFFYEI VEDDKSGRSS
DITSVRLEKISKCLEAESSDSGAESSEEEAAQE EVKGAEQSDND INDDHVEDVTGTQH L TSDSDSEVYCDSEMQFGQEESL
DSFTSNNGPFPQYYLGGHSSQPMENSGFREDI QVPPGNGNIGNMQVVAVEGKVEKVGKEDGRNNSGAPHRKRGGETDEF
SNVRRGGRHMRQHLSEGTGKRGYQSGSGDGERWGSGDRSGSLNEQIALVLMRLQEDMQNVLRQLRQLKLETLTAQKSSTS
TL0TAPSTQSRPSWPFEMSPGVLTFAI IWPFFIAQWLAYLYYQRRRRKL NLEG (SEQ ID NO: 42)

NOV21

In an alternative embodiment, a NOV2 variant is NOV21 of 1619 nucleotides (also referred to as CG51523-05_164718213), shown in Table 23A. A NOV21 variant differs from NOV2a by 5 amino acid changes, and an 11 amino acid insertion at position 161-162.

Table 23A. NOV21 nucleotide sequence (SEQ ID NO: 43).

AAGCTTCCACCATGTTCCAGTTTTCATGCAGGCTCTTGGGAAAGCTGGTCTGCTGCTGCTGCTGATTCCTCCGCCACAGGCCCT
TGGGACCGGGGCCAACACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGAGCGGCCGTGAA
GGTGATCCAGAGTTTGCCGAAGAATGGTTTCATTCCAGCCAACAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC
AGGCAACTGAAGGACCTCTGTAACCTTTCAAGGCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT
TCACTGGGTGATATGACCAAGAGGAGGCCAATATGTCATATGTTGAAGAATGAAAAGATTATTGAAACTATGCCAAAT
GACTGAGAAAGTTGAAGAATGCTCGCTGTCTATAGGCTCATTTTATGAAATTGTGCAGGACAAAAAGAGTGGCAGGAGTT
CTGATATAACCTCAGTCCGACTGGAGAAAATCTCTAAATGTTTGAAGATCTTGGTAATGTTCTCACTTCTACTCCGAAC
GCCAAAACCGTTAATGGTAAAGCTGAAGCAGTGCAGTGGAGCCGAGTCTGAGGAGAAGAGAGGCCCAAGAGAAGTGAA
AGGAGCAGAACAAAGTGATAATGATAAGAAAATGATGAAGAAGTCAGCAGACCAATAAGAATTGGAAGTCATTGTCATA
ATGGCTATGATAAAGATGGCTTTGTCAGGATATACAGAATGACATTCATGCCAGTTCTTCCCTGAATGGCAGAACCACT
GAAGAAGTAAAGCCCATTTGATGAAAACCTTGGGGCAAACTGGAATACTGCTGTTTGCATTACCAAGATATAAATGATGA
TCATGTTGAAGATGTTACAGGAATTCAGCATTTGACAAGCGATTTCAGACAGTGAAGTTTACTGTGATTCTATGGAACAAT
TTGGACAAGAAGAGTCTTTAGACAGCTTTACGTCCAACATGGACCATTTTCAGTATTACTTTGGGTGGTCTATCCAGTCAA
CCCATGGAATAATTCTGGATTCTGTGAAGATATTCAAGTACCTCCTGGAAATGGCAACATTTGGGAATATGCAGGTTGTGTC
AGTTGAAGGAAAAGCGAAGTCAAGCATGGAGGAGAAGATGGCAGGAATAACAGCGGAGCACACACCGGGAGGAGCGAG
CGCGAGAAACTGACGAATTTCTCTAATGTTAGAAGAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGGC
CGGCAGGTGGGAAGTGGAGGTGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCAGGCAGCCTCAATGAGCATATCGC
CCTCGTGCTGATGAGACTGCAGGAGGACATGCAGAATGTCCTTCAGAGACTGCAGAACTGGAACCGCTGACTGCTTTGC
AGGCAAAATCATCAACATCAACATTGCAGACTGCTCCTCAGCCCACCTCAGAGAGACCATCTTGGTGGCCCTTCGAGATG
TCTCCTGGTGTGCTAACGTTTGCCATCATATGCGCTTTTATTGCACAGTGTTGGTGATTTATACTATCAAAGAAGGAG
AAGAAAACCTGAACCTCGAG (SEQ ID NO: 43)

Table 23B. Encoded NOV21 protein sequence (SEQ ID NO: 44).

ASTMFQFHAGSWESWCCCCLI PADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQA
TEGPCKLSRPGFWDPIGRYKWDAWSSSLGDMTKEEAI IAYVEEMKKIIETMPMTEKVEELLRVIGFPFYEIVEDKKSGRSSI
ITSVRLEKISKCLEDLGNVLTSTPNKTVNGKAESSDSGAESEEEBAQEVEVGAEQSDNDKKMMKKSA DHKNLEVI VITNGY
DKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDSEMQFGQ
EESLDSFTSNNGPFPQYVLGGHSSQMGNSGFERIQVPPGNGNIGNOMQVVAEKGKVEKHGGEDGRNNSGAPHERERGGET
DEFSNVRRGRGHRMQLHSEGTKRGQVSGSGDGERWGSDRSGRSLNEHIALVLMRLQEDMQNVLRQLRQLKLETLALQAKS
STSLTOTAPORTSQRPSLWSPFEMS PGLVTFITAIWPFIAQWLVLVLYQRRRRKL NLE (SEQ ID NO: 44)

NOV2m

10 In an alternative embodiment, a NOV2 variant is NOV2m of 1619 nucleotides (also referred to as CG51523-05_166190452), shown in Table 24A. A NOV2m variant differs from NOV2a by 4 amino acid changes, and an 11 amino acid insertion at position 161-162.

Table 24A. NOV2m nucleotide sequence (SEQ ID NO: 45).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGTCTGCTGCTGCCTGATTCCCGCCGACAGACCT
TGGGACCGGGGCCAACACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGAGGCGGCCGTGAA
GGTGATCCAGAGTTTGCCGAAGAATGGTTTCATTCCAGCCAAACAAATGAAGTGATGCTTAAATTTTATAGCTTCTATAAGC
AGGCAACTGAAGGACCTGTAAACTTTCAAGGCCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTTGGAGT
TCATCGGGTGATATGACCAAGAGGAAGCCATGATGTCATATGTTGAAGAAATGAAAAGATTATTGAAACTATGCCAAT
CACTGAGAAAGTTGAAGAAATGCTCGCTGTTCATAGGTCATTTTATGAAATTGTTCGAGGACAAAAAGAGTGGCAGGAGTT
CTGATATAACCTCAGTCCGACTGGAGAAAATCTCTAAATGTTTAGAAGATCTTGGTAATGTTTCTCACTTCTACTCCAAAC
GCCAAAACCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAA
AGGAGCAGAACAAAGTGATATGATAAGAAAATGATGAAGAAGTCAGCAGACCATAAGAATTTGGAAGTCATTGTCTACTA
ATGGCTATGATAAAGATGGCTTTGTTCCAGGATATACAGAATGACATTTCATGCCAGTCTTCCCTGAAATGGCAGAACCT
GAAGAAGTAAAGCCTATTGTAGAAAACTTGGGGCAAACCTGGAAAATCTGCTGTTTGCATTACCAAGATATAAATGATGA
TCATGTTGAAGATGTTACAGGAATTGACATTTGACAAGCGATTGACACAGTGAAGTTTACTGTGATTCTATGGAACAAT
TTGGACAAGAAGAGTCTTTAGACAGCTTTACGTCCAACAATGGACATTTTTCAGTATTACTTGGGTTGGTCATTCCAGTCAA
CCCATTGAAAATTTCTGGATTTCTGTGAAGATATTCAAGTACCTCCTGGAAATGGCAACTTGGGAATATGGAGGTGGTTGCG
AGTTGAAGGAAAGGTGAAGTCAGCATGGAGGAGGAAGATGGCAGGAATAACAGCGGAGCGCCACACCGGGAGAAGCGAG
CGCGAGAAACTGATGAATTTCTCTAATGTTAGAAGAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGGC
CGGCAGGTGGGAAGTGGAGATGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGC
CCTCGTGCTGATGAGACTGACAGGAGACATGCAGAATGTCCTTCAGAGACTGCAGAAACTGGAACCGCTGACTGCTTTGCG
AGGCAAAATCATCAACATCAACATTGCAGACTGCTCCTCAGCCCACCTCAGAGACCATTTGGTGGCCCTTCGAGATG
TCTCTGGTGTGCTAACGTTTGCCATCATATGGCCCTTTTATGTCACAGTGGTGGTGTATTATACTATCAAAGAAGGAG
AAGAAAACCTGAACCTCGAG (SEQ ID NO: 45)

Table 24B. Encoded NOV2m protein sequence (SEQ ID NO: 46).

ASTMFQFHAGSWESWCCCLIPADRPDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPKNGSFQPTNEVMLKFYSFYKQA
TEGPCKLSRPGFWDPIGRYKWDAWSSSLGDMTKEEAMIAYVEEMKKIIETMPMTEKVEELLRVIGPFYEIVEDKKSGRSSD
ITSVRLEKISKLEDLGNVLSTSTPNAKTVNGKAESSDGAEESEEAQEEVKGAEQSDNDKMKMKKSADHKNLELVITNG
YDKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDMEQFG
QEESLDSFTSNNGPFPYYLGGHSSQPMENSGFREDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRG
ETDEFSNVRRGRGHRMQHLSEGTKGRQVGSGDDGERWGSDRGSRGSLNEQIALVLMRLQEDMQNVLQRLQKLETLTALQA
KSSTSTLTQAPQPTSQRPSWWPFEMSPGVLTFAIIWPFIAQWLVLVLYYQRRRRKLNLE (SEQ ID NO: 46)

5 In an alternative embodiment, a NOV2 variant is NOV2n of 1619 nucleotides (also referred to as CG51523-05_166190467), shown in Table 25A. Similarly to a NOV2n variant, a NOV2n variant differs from NOV2a by 4 amino acid changes, and an 11 amino acid insertion at position 161-162.

Table 25A. NOV2n nucleotide sequence (SEQ ID NO: 47).

AAGCTTCCACCATGTTCCAGTTCATGCAAGGCTCTTGGGAAAGCTGGTGCTGCTGCTGCCTGATTCCC GCCGACAGACCT TGGGACCGGGGCCAACACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGTTTT GAGCGCGCCGTGAA GGTGATCCAGAGTTTGGCCGAAGAATGGTTTCATTCCAGCCAAACAAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC AGGCAACTGAAGGACCTGTAAACCTTTCAAGCCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT TCATCTGGGTGATATGACCAAAGAGGAAGCCATGATGCATATGTTGGAAGAAATGAAAAGATTATTGAACTATGCCAAT GACTGAGAAAGTTGAAGAATTGCTGCGTGT CATAGGTCCATTTTATGAAATTGTCGAGGACAAAAGAGTGGCAGGAGTT CTGATATAACCTCAGTCCGACTGGAGAAAATCTCTAAATCTTTAGAAAGATCTTGGTAAATGTTCTCACTTCTACTCCAAAC GCCAAAACCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGCTCTGAGGAAGAAGAGGCCCAAGAAGAGTGAA AGGACGAGAACAAAGTGATAATTGATAAGAAATGATGAAGAAGTCAGCAGACATAAGAATTTGGAAGTCATTGTCTACTA ATGGCTATGATAAAGATGGCTTTGTTCAGGATATGCAGAATGACATTTCATGCCAGTTCCTCCCTGAATGGCAGAAGCACT GAAGAAGTAAGGCCTATTGATGAAAACCTGGGGCAAACCTGGAAAATCTGCTGTTTGCATTTCAACCAAGATATAAATGACGA TCATGTTGAAGATGTTACAGGAATTCAGCATTTGACAAGCGATTTCAGACAGTGAAGTTTACTGTTCTTATGGAACAA TTGGACAAGAAGACTCTTTAGACAGCTTTACGTCCAACAATGGACCATTTTCAGTATTACTTGGTGTTGTCATTCCAGTCAA CCCATGGAAAATCTCGATTTCTGTGAAGATATTCAAGTACCTCCTGGAAAATGGCAACATTGGGAATATGCAGGTGGTTGC

Table 25B. Encoded NOV2n protein sequence (SEQ ID NO: 48).

ASTMFQFHAGSWESWCCCLIPADRPDRGQHWQLEMADTRSVHETRFEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQA
TEGPCKLSRPGFWDPIGRYKWDASSLGDMTKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYIEIVEDKSGRSSD
ITSVRLEKISKCLEDLGNVLTSTPNAKTVNGKAESSDSGAESSEEEAAQEEVKGAEQSDNDKKMMKKSADHKNLEIVITNG
YDKDGFVQDMQNDIHASSSLNGRSTEEVRPIDENLGQTGKSAVCIHQDINDDHVEDVTGIIQHLTSDSDSEVYCDSEMGFG
QEESLDSFTSNNGPFQYYLGGHSSQPMENSGFREDIQVPPGNGNIGNMQVVAVEGKEGVKHGIGEDGRNNSGAPHREKRG
ETDEFSNVRRGRRGRMQHLSGEGTKGRQVSGGDDGERWSDRSGRSLNEQIALVLMRLQEDMQNVLQRLQKLETLTALQA
KSSTSTLQTAPQPTSQRPSWWPFEMSPGVLTFAIWPFIQWLVLVYLYYQRRRRKLNLE (SEQ ID NO: 48)

In an alternative embodiment, a NOV2 variant is NOV2o of 1619 nucleotides (also referred to as CG51523-05_166190475), shown in Table 26A. A NOV2o variant differs from NOV2a by 3 amino acid changes at positions 170, 372 and 403, and an 11 amino acid insertion at position 161-162.

AAGCTTCCACCATGTTCCAGTTTCATGCAAGGCTCTTGGGAAAGCTGGTGTCTGCTGCTGCCTGATTCCCGCCGACAGACCTTGGGACCGGGCCAACTGCGAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGAGCGCGCGGTGAAAGGTGATCCAGAGTTTGCCGAAGAATGGTTTCATTCCAGCCAACAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGCAGGCAACTGAAGGACCTCTGAATCTTCAAGGCCCTGGATTGTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGTTCACTGGGTGATATGACCAAGAGGAAGCCATGATGTCATATGTTGAAGAAATGAAAAAGATTATTGAAACTATGCGCAATGACTGAGAAAGTTGAAGAATTGCTGCGTGTCTAGGTCCATTTTATGAAATTGTGCGAGGACAAAAAGAGTGGCAGGAGTCTGATATAACCTCAGTCCGACTGGAGAAAAATCTCTAAATGTTTAGAAGATCTTGGTAATGTTCTCACTTCTACTCCGAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAAAGGAGCAGAACAAAGTGATAATGATAAGAAAAATGATGAAGAAGTCAGCAGACCATAAGAATTGGAAGTCATTGTCTACTATGGCTATGATAAAGATGGCTTTGTTCAGGATATACAGAATGACATTTCATGCCAGTTCTCCCTGAATGCGCAGAAGCACTGAAGAAGATAAAGCCCATTTGATGAAAACCTTGGGGCAAACCTGGAAAACTGCTGTTTGCATTACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTGAGCATTGTGACAAGCGATTGACAGAGTGAAGTTTACTGTGATTCTATGGAACAATTTGGACAAGAAGAGTCTTTAGACAGCTTTACGTCCAACAATGGAACCATTTGATTAATTACTTGGGTGGTCAATCCAGTCAACCCATGGAAAAATCTGGATTTCGTGAAGATATTCAGAGTACCTCTGGAAATGGCAACATTGGGAATATGCAAGAGTGGTTGCAGTTGAAGGAAAAGGTGAAGTCAGCATGAGGAGGAAGAGGCGAGGAGTAACACGCGAGCACCCGGGAGAGCGGAGCGGAGAAACTGACGAATTCTCTAATGTTAGAAGAGGAAGAGGACATAGGATGCAACACCTGAGCGAAGGAACCAAGGGCCGGCAGGTGGGAAGTGGAGGTGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGCCCTCGTGCTGATGAGACTGCGAGAGGACATGCGAATGTCTTCAGAGACTGCGAAACTGGAAACCGTGACTGCTTTGCAAGGCAAAATCATCAACATCAACATTGCGAGACTGCTCCTCAGCCCACCTCACAGAGACCATTGGTGGCCCTTCGAGATGTCTCTGGTGTGCTAATACGTTGGCATCATATGGCCTTTATTGACAGATGGTGGTGGTGTATTATATACTATCAAGAAGGAGAGAAGAAACTGAACCTCGAG (SEQ ID NO: 49)

ASTMFQFHAGSWESWCCCCLI PADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQ
ATEGPCKLSRPGFWDPIGRYKWDAWSSLGDMTKEEAMIAIVEEMKKI IETMPMTEKVEELLRVIGPFYEIVEDKKSGRSS
DITSVRLEKISKCEDLGNVLSTPNAKTVNGKAESSDSGAEESEEEAAIEVKGAEQSDNDKMMKKKSAHKNLEIVTN
GYDKGFVQDIQNDIHASSNLNGRSTEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDSEMQF
GQEESLDSFTSNNGPFFQYYLGGHSSQPMENSGFREDIQVPPGNGNIGNMQVVAVEGKGEVKHGGEGRNNSGAPHREKRG
GETDEFSNVRRGRGHRMQLSEGTKRQVSGSGDGERWGSDRGSRGSLNEQIALVLMRLQEDMQNLQRLQKLETLTALQ
AKSSTSTLQTAPQTSQRPSWPFEMSPGVLTFAIIWPFIAQWLVLVYLYQRRRRKLNLE (SEQ ID NO: 50)

NOV2p

In an alternative embodiment, a NOV2 variant is NOV2p of 1619 nucleotides (also referred to as CG51523-05_166190498), shown in Table 27A. A NOV2p variant differs from NOV2a by 2 amino acid changes at positions 170 and 403, and an 11 amino acid insertion at position 161-162.

Table 27A. NOV2p nucleotide sequence (SEQ ID NO: 51).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGCTGCTGCTGCCTGATTCCCGCCGACAGACCT TGGGACCGGGCCAACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGAGGCGGCCGTGAA GGTGATCCAGAGTTTGCCGAAGAATGGTTCATTCCAGCCAACAAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC AGGCAACTGAAGGACCTGTAACTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT TCACTGGGTGATATGACCAAAGAGGAAGCCATGATTGCATATGTTGAAGAAATGAAAAAGATTATTGAAACTATGCCAAT GACTGAGAAAGTTGAAGAATTGCTGCGTGTCTATAGGTCCATTTTATGAAATTGTCTGAGGACAAAAAGAGTGGCAGGAGTT CTGATATAACCTCAGTCCGACTGGAGAAAACTCTCTAAATGTTTAGAAGATCTTGGTAATGTTCTCACTTCTACTCCGAAC GCCAAAACCGTTAATGGTAAAGCTGAAAGCAGTGACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAA AGGAGCAGAACAAAGTGATAATGATAAGAAATGATGAAGAAGTCAGCAGACCATAAGAATTTGGAAGTCATGTCTACTA ATGGCTATGATAAAGATGGCTTTGTTCAGGATATACAGAATGACATTTCATGCCAGTTCTTCCCTGAATGGCAGAAGCACT GAAGAAGTAAAGCCATTGATGAAAATCTGGGGCAAACCTGGAAAATCTGCTGTTTGCAATTCACCAAGATATAAATGATGA TCATGTTGAAGATGTTACAGGAATTCAGCATTTCAGCAAGCGATTTCAGACAGTGAAGTTTACTGTGATTCTATGGAACAAT TTGGACAAGAAGAGTCTTTAGACAGCTTTACGTCCAACAATGGACCATTTCAGTATTACTTGGGTGGTTCATTCAGTCAA CCCATGGAAAATCTGGATTTCGTGAAGATATTCAAGTACCTCTGGAAATGGCAACATTGGGAATATGCAGGTGGTTGC AGTTGAAGGAAAAGGTGAAGTCAAGCATGGAGGAGAAGATGGCAGGAATAACAGCGGAGCACCACACCGGGGAGAAGCGAG GCGGAGAACTGACGAATTCCTAATGTTAGAAGAGGAAGAGGACATAGGATGCAACACTTGAGCGAAGGAACCAAGGC CGGCAGGTGGGAAGTGGAGGTGATGGGGAGCGCTGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGC CCTCGTGCTGATGAGACTGCAGGAGGACATGCAGAATGTCCTTCAGAGACTGCAGAACTGGAAACGCTGACTGCTTTGC AGGCAAAATCATCAACATCAACATTGCAGACTGCTCCTCAGCCACCTCACAGAGACCATCTTGGTGGCCCTTCGAGATG TCTCCTGGTGTGCTAACGTTTGCCATCATATGGCCTTTTATTGCACAGTGGTTGGTGTATTATATACTATCAAAGAAGGAG AAGAAAATGAACCTCGAG (SEQ ID NO: 51)
--

Table 27B. Encoded NOV2p protein sequence (SEQ ID NO: 52).

ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQ ATEGPCKLSRPGFWDPIGRYKWDWSSLDGMTKEEAMIAVVEEMKKIIETMPMTEKVEELLRVIGPFYFIVEDKKSGRSS DITSVRLEKISKLEDLGNVLTSTPNAKTVNGKAESSDSGAEEEEAEVVKGAEQSDNDKMMKKSAADHKNLEIVTN GYDKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDSEMEQF GQEESLDSFTSNNGFPQYYLGGHSSQPMENSGFREDIQVPPNGNIGNMQVVAVEGKGEVKGHGGEDGRNNSGAPHREKRG GETDEFSNVRRGRGRMQHLSGKTGRQVGSAGDGERWGSRRGSLNEQIALVLMRLQEDMQNVLQRLQKLETLTALQ AKSSTSTLQTAPQTSQRPSWWPFEMSPGVLTFAIWPFIQWLVLVYQRRRRKLNLE (SEQ ID NO: 52)
--

NOV2q

In an alternative embodiment, a NOV2 variant is NOV2q of 1586 nucleotides (also referred to as CG51523-05_166190460), shown in Table 28A. A NOV2q variant differs from NOV2a by 3 amino acid changes at positions 170, 231 and 463.

Table 28A. NOV2q nucleotide sequence (SEQ ID NO: 53).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGCTGCTGCTGCCTGATTCCCGCCGACAGACCT TGGGACCGGGCCAACTGGCAGCTGGAGATGGCGGACACGAGATCCGTGCACGAGACTAGGTTTGAGGCGGCCGTGAA GGTGATCCAGAGTTTGCCGAAGAATGGTTCATTCCAGCCAACAAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC

AGGCAACTGAAGGACCTGTAACTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT
TCACTGGGTGATATGACCAAAGAGGAAGCCATGATTGCATATGTTGAAGAAATGAAAAAGATTATTGAAACTATGCCAAT
GACTGAGAAAGTTGAAGAATTGCTGCGTGTCTATAGGTCCATTTTATGAAATTGTTCGAGGACAAAAAGAGTGGCAGGAGTT
CTGATATAACCTCAGATCTTGGTAATGTTCTCACTTCTACTCCAAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGT
GACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAAGAGGACAGAACAAGTGATAATGATAAGAAAAT
GATGAAGAAGTCAGCAGACCATAAGAATTTGGAAGTCATTGTCTAATGGCTATGATAAAAATGGCTTTGTTCAGGATA
TACAGAATGACATTATGCCAGTTCTTCCCTGAATGGCAGAAGCACTGAAGAAGTAAAGCCCATGATGAAAACCTTGGGG
CAAACCTGGAAAATCTGCTGTTTGCATTACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTGAGCATT
GACAAGCGATTGACAGAGTGAAGTTTACTGTGATTCTATGGAACAATTGGACAAGAAGAGTCTTTAGACAGCTTTACGT
CCAACAATGGACCATTTTCAGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAATCTGGATTTCGTGAAGATATT
CAAGTACCTCCTGGAAATGGCAACATTGGGAATATGCAGGTGGTTGCAGTTGAAGGAAAAGGTGAAGTCAAGCATGGAGG
AGAAGATGGCAGGAATAACAGCGGAGCACACACCGGGAGAAGCGAGGCGGAGAACTGACGAATTCTCTAATGTTAGAA
GAGGAAGAGGACATAGGATGCAACACTTGAAGCAAGGAACCAAGGGCCGCGAGGTGGGAAGTGGAGGTGATGGGGAGCGC
TGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGCCCTCGTGTGATGAGACTGCAGGAGGACATGCA
GAATGTCCTTCAGAGACTGCAGAACTGGAAACGCTGACTGCTTTGCAGGCAAAATCATCAACATCAACATTGCAGACTG
CTCCTCAGCCCACTCAGAGACCATTCTGGTGGCCCTTCGAGATGTCTCCTGGTGTGCTAACGTTTGCCATCATATGG
CCTTTTATTGCACAGTGGTTGGTGTATTATATACTATCAAGAAGGAGAAGAAAACCTGAACCTCGAG (SEQ ID NO:
53)

Table 28B. Encoded NOV2q protein sequence (SEQ ID NO: 54).

ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEAQVKVIQSLPKNGSFQPTNEMMLKFYSFYKQA
TEGPCKLSRPGFWDPIGRYKWDAWSSLDGMTKEEAMIAVVEEMKKIETMPMTEKVEELLRVIGPFYEIVEDKKSGRSSDI
TSDLGNVLTSTPNAKTVNGKAESSDGAEEEEEAEVKGAEQSDNDKMMKKSADHKNLEVIIVTNGYDKNGFVQDIQND
IHASSSLNGRSTEEVKPIDENLGGTQKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDMSMEQFQGEESLDSFTSNNGP
FQYYLGGHSSQPMENSGFREDIQVPPGNGNIGNMQVVAVEGKGEVKGHGEDGRNNSGAPHREKRGGETDEFNSVRRRGHR
MQHLSEGTGKRQVGSAGGGERWGSDRGSRGSLNEQIALVLMRLQEDMQNVLRQLKLETLTALQAKSSTSTLQAPQPTSQ
RPSWWPFEMSPGVLTFAIWPFFIAQWLVLVYQRRRRKLNLE (SEQ ID NO: 54)

NOV2r

In an alternative embodiment, a NOV2 variant is NOV2r of 1586 nucleotides (also
5 referred to as CG51523-05_166190483), shown in Table 29A. A NOV2r variant differs from
NOV2a by 5 amino acid changes at positions 170, 342, 396, 403, and 452.

Table 29A. NOV2r nucleotide sequence (SEQ ID NO: 55).

AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGGGAAAGCTGGTGTCTGCTGCTGCTGCTGATTCCCGCCGACAGACCT
TGGGACCGGGCCCAACATGGCAGCTGGAGATGGCGGACACAGATCCGTGCACGAGACTAGGTTTGGAGCGGCGGTGAA
GGTGATCCAGAGTTTGGCGAAGAATGGTTCATTCCAGCCAACAATGAAATGATGCTTAAATTTTATAGCTTCTATAAGC
AGGCAACTGAAGGACCTGTAACTTTCAAGGCCTGGATTTTGGGATCCTATTGGAAGATATAAATGGGATGCTTGGAGT
TCACTGGGTGATATGACCAAAGAGGAAGCCATGATTGCATATGTTGAAGAAATGAAAAAGATTATTGAAACTATGCCAAT
GACTGAGAAAGTTGAAGAATTGCTGCGTGTCTATAGGTCCATTTTATGAAATTGTGCAAGACAAAAAGAGTGGCAGGAGTT
CTGATATAACCTCAGATCTTGGTAATGTTCTCACTTCTACTCCGAACGCCAAAACCGTTAATGGTAAAGCTGAAAGCAGT
GACAGTGGAGCCGAGTCTGAGGAAGAAGAGGCCCAAGAAGAAGTGAAGGAGCAGAACAAGTGATAATGATAAGAAAAT
GATGAAGAAGTCAGCAGACCATAAGAATTTGGAAGTCATTGTCTAATGGCTATGATAAAGATGGCTTTGTTTCAGGATA
TACAGAATGACATTATGCCAGTTCTTCCCTGAATGGCAGAAGCACTGAAGAAGTAAAGCCCATGATGAAAACCTTGGGG
CAAACCTGGAAAATCTGCTGTTTGCATTACCAAGATATAAATGATGATCATGTTGAAGATGTTACAGGAATTCAGCATT
GACAAGCGATTGACAGAGTGAAGTTTACTGTGATTCTATGGAACAATTGGACAAGAAGAGTCTTTAGACAGCTTTACGT
CCAACAATGGACCATTTTCAGTATTACTTGGGTGGTCATTCCAGTCAACCCATGGAAAATCTGGATTTCGTGAATATATT
CAAGTACCTCCTGGAAATGGCAACATTGGGAATATGCAGGTGGTTGCAGTTGAAGGAAAAGGTGAAGTCAAGCATGGAGG
AGAAGATGGCAGGAATAACAGCGGAGCACACACCGGGAGAAGCGAGGCGGAGAACTGACGAATTCTCTAATGTTGGAA
GAGGAAGAGGACATAGGATGCAACACTTGAAGCAAGGAACCAAGGGCCGCGAGGTGGGAAGTGGAGGTGATGGGGAGCGC
TGGGGCTCCGACAGAGGGTCCCGAGGCAGCCTCAATGAGCAGATCGCCCTCGTGTGATGAGACTGCAGGAGGACATGCA
GAATGTCCTTCAGAGACTGCAGAACTGGAAACGCCGACTGCTTTGCAGGCAAAATCATCAACATCAACATTGCAGACTG
CTCCTCAGCCCACTCAGAGACCATTCTGGTGGCCCTTCGAGATGTCTCCTGGTGTGCTAACGTTTGCCATCATATGG
CCTTTTATTGCACAGTGGTTGGTGTATTATATACTATCAAGAAGGAGAAGAAAACCTGAACCTCGAG (SEQ ID NO:
55)

Table 29B. Encoded NOV2r protein sequence (SEQ ID NO: 56).

ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPKNGSFQPTNEMMLKFYSFYKQ
ATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIYVEEMKKIETMPMTEKVEELLRVIGPFYEIVEDKKSGRSS
DITSDLGNVLTSTPNAKTVNGKAESSDGAEESEEEAAQEEVKGAEQSDNDKMMKKKSADHKNLEIVITNGYDKDGFVQDI
QNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQDINDDHVEDVTGIQHLTSDSDSEVYCDSEMQFGQEESELDSTFS
NNGPFQYYLGGHSSQPMENSGFREYIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFENVGR
GRGHRMQHLESGTGKRGVSGDGERWGS DRGSRGSLNEQIALVLMRLQEDMQNVLRQLQKLETPTALQAKSSTSTLQTA
PQPTSQRPSWWPFEMSPGVLTFAIWPFIQWLVLVYQRRRRKLNLE (SEQ ID NO: 56)

The homology between NOV2a and its variants is shown graphically in the ClustalW analysis shown in Table 30. In the ClustalW alignment of NOV2a protein and its variants, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 30: Clustal W of NOV2 variants

NOV2a	---MFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	57
NOV2b	-----KLD RPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	39
NOV2c	-----KLT RFEEAAVKVIQSLPK	17
NOV2d	-----KLT RFEEAAVKVIQSLPK	17
NOV2e	-----KLT RFEEAAVKVIQSLPK	17
NOV2f	-----KLT RFEEAAVKVIQSLPK	17
NOV2g	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2h	GFHHVPVSCRLLGKLVLLLPDSADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2i	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2j	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2k	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2l	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2m	ASTMFQFHAGSWESWCCCLIPADRP-----DRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	59
NOV2n	ASTMFQFHAGSWESWCCCLIPADRP-----DRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	59
NOV2o	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2p	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2q	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2r	ASTMFQFHAGSWESWCCCLIPADRPWDRGQHWQLEMADTRSVHETRFEEAAVKVIQSLPK	60
NOV2a	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	117
NOV2b	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	99
NOV2c	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	77
NOV2d	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	77
NOV2e	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	77
NOV2f	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	77
NOV2g	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	120
NOV2h	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	120
NOV2i	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	120
NOV2j	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	120
NOV2k	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	120
NOV2l	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	120
NOV2m	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	119
NOV2n	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	119
NOV2o	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	120
NOV2p	NGSFQPTNEMMLKFYSFYKQATEGPKCLSRPGFWDPIGRYKWDWSSSLGDMTKEEAMIAIY	120

NOV2q **N**SSFQPT**N**EMMLKFPYSFYKQATEGPKLSRPGFWDPIGRYKWDASSLGDMTKEEAMIA**V** 120
NOV2r **N**SSFQPT**N**EMMLKFPYSFYKQATEGPKLSRPGFWDPIGRYKWDASSLGDMTKEEAMIA**V** 120

5
NOV2a **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**-----**D**LGN**V** 166
NOV2b **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**-----**D**LGN**V** 148
NOV2c **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 137
NOV2d **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 137
NOV2e **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 134
10
NOV2f **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**-----**D**LGN**V** 126
NOV2g **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**-----**D**LGN**V** 169
NOV2h **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 180
NOV2i **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**-----**D**LGN**V** 169
NOV2j **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**-----**D**LGN**V** 169
NOV2k **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VR-----**L**EK**S**K**C** 173
15
NOV2l **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 180
NOV2m **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 179
NOV2n **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 179
NOV2o **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 180
NOV2p **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**VRLEKISKCL**E****D**LGN**V** 180
20
NOV2q **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**-----**D**LGN**V** 169
NOV2r **V**EEMKKI**I**ETMPMT**E**KVEELLRVIGPFY**E**IVEDKKSGRSSD**I**T**S**-----**D**LGN**V** 169

25
NOV2a **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 226
NOV2b **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 208
NOV2c **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 197
NOV2d **L**TST**P****N**AKTV**G**KAESS**D**SGAE**S**EE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 196
NOV2e -----**E**EE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 171
NOV2f **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVK**G**-----**E**QSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 185
NOV2g **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 229
30
NOV2h **L**TST**P****N**AKTV**G**KA**E**SS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 240
NOV2i **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 229
NOV2j **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 229
NOV2k **L**-----**E**ASS**D**SGAESEEE**A**QEEVKGAEQSD**N**----- 203
35
NOV2l **L**TST**P****N**-----**K**TV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 239
NOV2m **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 239
NOV2n **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 239
NOV2o **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 240
NOV2p **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 240
NOV2q **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 229
40
NOV2r **L**TST**P****N**AKTV**G**KAESS**D**SGAESEEE**A**QEEVKGAEQSD**N**DKMMK**S**ADH**K**NLE**V**IV**T**N 229

45
NOV2a **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 286
NOV2b **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 268
NOV2c **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 257
NOV2d **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 256
NOV2e **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 231
NOV2f **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 244
NOV2g **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 289
NOV2h **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 300
50
NOV2i **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 289
NOV2j **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 289
NOV2k -----**I**N**D**DH**V**ED**V**T**G** 214
NOV2l **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 299
NOV2m **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 299
55
NOV2n **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 299
NOV2o **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 300
NOV2p **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 300
NOV2q **G**YDK**K****G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 289
60
NOV2r **G**YDKD**G**FVQD**I**Q**N**DIHASS**L**NGRST**E**EV**K**PID**E**N**L**GQ**T**GKS**A**VC**I**HQD**I**N**D**DH**V**ED**V**T**G** 289

5

NOV2a **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 346
 NOV2b **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 328
 NOV2c **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 317
 NOV2d **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 316
 NOV2e **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 291
 NOV2f **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 304
 NOV2g **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 349
 NOV2h **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 360
 NOV2i **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 349
 NOV2j **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 349
 NOV2k **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 274
 NOV2l **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 358
 NOV2m **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 359
 NOV2n **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 359
 NOV2o **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 360
 NOV2p **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 360
 NOV2q **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 349
 NOV2r **IQHLTSDSDSEVYCDSEMFQGGQESLDSFTSNNGPFQYVLGGHSSQPMENSGFREDIQVF** 349

20

NOV2a **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 406
 NOV2b **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 388
 NOV2c **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 376
 NOV2d **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 375
 NOV2e **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 350
 NOV2f **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 363
 NOV2g **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 409
 NOV2h **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 420
 NOV2i **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 409
 NOV2j **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 386
 NOV2k **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 334
 NOV2l **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 418
 NOV2m **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 419
 NOV2n **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 419
 NOV2o **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 420
 NOV2p **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 420
 NOV2q **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 409
 NOV2r **PCNENIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRGRGHRMQL** 409

35

40

NOV2a **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 466
 NOV2b **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 448
 NOV2c **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 436
 NOV2d **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 435
 NOV2e **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 410
 NOV2f **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 423
 NOV2g **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 469
 NOV2h **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 480
 NOV2i **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 469
 NOV2j **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 446
 NOV2k **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 394
 NOV2l **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 478
 NOV2m **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 479
 NOV2n **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 479
 NOV2o **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 480
 NOV2p **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 480
 NOV2q **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 469
 NOV2r **SEGTKGRQVSGGDDGERWGS DRGSRGSLNEQIALVLMRLQEDMCNVLQRLQKLETLTALQ** 469

50

55

60

NOV2a **AKSSTSTLTQAPQTSQRPSWPFEMSPGVLTFAIWPPIAQWLVLVYQRRRRKIN---** 523 (SEQ ID NO: 22)
 NOV2b **AKSSTSTLTQAPQTSQRPSWPFEMSPSR-----** 477 (SEQ ID NO: 24)
 NOV2c **AKSSTSTLTQAPQTSQRPSWPFEMSP-----LE-** 466 (SEQ ID NO: 26)

5	NOV2d	AKSSTSTLQTAPQPTSQRPSWPPFEMSP-----LE-	465 (SEQ ID NO: 28)
	NOV2e	AKSSTSTLQTAPQPTSQRPSWPPFEMSP-----LE-	440 (SEQ ID NO: 30)
	NOV2f	AKSSTSTLQTAPQPTSQRPSWPPFEMSP-----LE-	453 (SEQ ID NO: 32)
	NOV2g	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	528 (SEQ ID NO: 34)
	NOV2h	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	539 (SEQ ID NO: 36)
	NOV2i	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	528 (SEQ ID NO: 38)
	NOV2j	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	505 (SEQ ID NO: 40)
	NOV2k	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	454 (SEQ ID NO: 42)
10	NOV2l	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	537 (SEQ ID NO: 44)
	NOV2m	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	538 (SEQ ID NO: 46)
	NOV2n	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	538 (SEQ ID NO: 48)
	NOV2o	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	539 (SEQ ID NO: 50)
	NOV2p	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	539 (SEQ ID NO: 52)
	NOV2q	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	528 (SEQ ID NO: 54)
15	NOV2r	AKSSTSTLQTAPQPTSQRPSWPPFEMSPGVLTFAIWPFIAQWLVLVYQRRRRKLNLE-	528 (SEQ ID NO: 56)

Based on the relatedness between NOV2a, its variants, and endozepine-related protein precursor-like proteins, the disclosed NOVX proteins are novel members of the endozepine-related protein precursor-like protein family. Therefore, the nucleic acids and proteins of the inventions are useful in potential therapeutic applications implicated in various pathologies and disorders described and other pathologies and disorders related to aberrant function or aberrant expression of these endozepine-related protein precursor-like proteins.

Acyl-CoA-binding protein (ACBP) is a small (10 Kd) protein that binds medium- and long-chain acyl-CoA esters with high affinity, and may act as an intra-cellular carrier of acyl-CoA esters. ACBP is a highly conserved protein that has been so far found in vertebrates, insects, plants and yeast. ACBP has a number of important physiological and biochemical functions: it is known as a diazepam binding inhibitor, as a putative neurotransmitter, as a regulator of insulin release from pancreatic cells, and as a mediator in corticotropin-dependent adrenal steroidogenesis. It is possible that the protein acts as a neuropeptide that takes part in the modulation of gamma-aminobutyric acid-ergic transmission. The structure of ACBP has been deduced by NMR spectroscopy and has been shown to be a mainly-alpha protein, consisting of 5 short alpha-helices and 3 connecting beta-strands.

Other proteins belonging to the ACBP family include mouse endozepine-like peptide (ELP); mammalian MA-DBI, a transmembrane protein of unknown function which has been found in mammals (like the one reported here in this invention); and human DRS-1, a protein of unknown function that contains a N-terminal ACBP-like domain and a C-terminal enoyl-CoA isomerase/hydratase domain.

Benzodiazepines modulate signal transduction at type A GABA (gamma-aminobutyric acid) receptors located in brain synapses. GABA is the predominant inhibitory neurotransmitter of the mammalian central nervous system. This receptor binds GABA, beta-carbolines, and benzodiazepines with high affinity and a chloride ion channel.

Benzodiazepines prolong the chloride ion channel opening burst elicited by GABA and thereby enhance GABA-mediated inhibitory responses. This facilitation plays a role in reducing pathologic anxiety. An endogenous ligand has been identified that is recognized by the beta-carboline/benzodiazepine recognition site located in the GABA receptor. This ligand, diazepam binding inhibitor (DBI), is a protein of about 11 kD that displaces beta-carbolines and benzodiazepines bound to brain membrane fractions in vitro. DBI or a derivative small neuropeptide is thought to downregulate the effects of GABA.

A polypeptide related to DBI, with similar binding activity to diazepam, has been isolated from human and bovine brain. This protein, called endozepine, contains 86 amino residues. Northern analysis using the cloned cDNA demonstrated that the message is expressed in heart, liver, and spleen, in addition to brain. Benzodiazepine receptors unassociated with the GABA receptor complex, and distinct from those seen in association with the central nervous system, have been identified in peripheral tissues.

Bovine and human cDNA sequences encoding a putative benzodiazepine receptor ligand. cDNAs containing the entire coding sequence of endozepine, a putative ligand of the benzodiazepine receptor, were isolated from bovine and human cDNA libraries. These libraries were constructed using a novel oligonucleotide adapter molecule that allowed us to combine the use of G/C tailing with the preservation of the unique Eco RI site in the vector, lambda gt10. The amino acid sequences derived from these cDNA clones are identical to those previously determined for the purified proteins and are homologous to a related rat protein termed diazepam-binding inhibitor. The endozepine proteins are highly conserved, as illustrated by the finding that the nucleotide sequences of the coding regions are 93% conserved between the bovine and human forms. Analysis of these sequences indicates that endozepine is not, as expected, derived from a precursor molecule containing a transient signal peptide. Moreover, Northern analyses using the cloned cDNAs as hybridization probes indicate that the 650-nucleotide endozepine mRNA is expressed in a number of peripheral tissues in addition to brain. These observations may be consistent with a recent report describing the presence in peripheral tissues of benzodiazepine receptors on the outer mitochondrial membrane (Anholt et al., 1986). In addition to the endozepine cDNAs, a bovine cDNA clone was isolated that encodes a larger protein, a portion of which is homologous to endozepine. This related protein may be synthesized in a precursor form containing putative signal peptide and membrane-spanning domains.

Human endozepine, an 86 amino acid polypeptide, was originally isolated from human brain tissue as a putative ligand of the benzodiazepine receptor. Complete amino acid sequencing of the human and bovine proteins revealed significant homology with the partial sequence of diazepam binding inhibitor (DBI), a protein from rat brain. Both endozepine and DBI have been shown to elicit behavioral effects, suggesting that they function as pharmacologically-active ligands of the GABA (gamma-aminobutyric acid) receptor complex. Subsequent cDNA cloning of human and bovine endozepine, rat DBI and human DBI has shown that these proteins are encoded by the same gene. A related cDNA, encoding a transmembrane protein of 533 amino acids with a domain homologous to DBI, has also been cloned from bovine brain.

Diazepam binding inhibitor (DBI) is a 10-kDa polypeptide that regulates mitochondrial steroidogenesis, glucose-induced insulin secretion, metabolism of acyl-CoA esters, and the action of gamma-aminobutyrate on GABAA receptors. To investigate the regulation of DBI gene expression, three positive clones were isolated from a rat genomic library. One of them contained a DBI genomic DNA fragment encompassing 4 kb of the 5' untranslated region, the first two exons, and part of the second intron of the DBI gene. Two other overlapping clones contained a processed DBI pseudogene. Several transcription initiation sites were detected by RNase protection and primer extension assays. Different tissues exhibited clear differences in the efficiencies of transcription startpoint usage. Transient expression experiments using DNA fragments of different length from the 5' untranslated region of the DBI gene showed that basal promoter activity required 146 bp of the proximal DBI sequence, whereas full activation was achieved with 423 bp of the 5' untranslated region. DNase I protection experiments with liver nuclear proteins demonstrated three protected regions at nt -387 to -333, -295 to -271, and -176 to -139 relative to the ATG initiation codon; in other tissues the pattern of protection was different. In gel shift assays the most proximal region (-176 to -139) was found to bind several general transcription factors as well as cell type-restricted nuclear proteins which may be related to specific regulatory patterns in different tissues. Thus, the DBI gene possesses some features of a housekeeping gene but also includes a variable regulation which appears to change with the function that it subserves in different cell types.

30

A trypsin-sensitive cholecystokinin-releasing peptide (CCK-RP) has been isolated from porcine and rat intestinal mucosa. The amino acid sequence of this peptide was determined to be identical to that of the diazepam-binding inhibitor (DBI). To test the role of DBI in pancreatic secretion and responses to feeding, pancreaticobiliary and intestinal cannula

were used to divert bile-pancreatic juice from anesthetized rats. Within 2 hours, this treatment caused a 2-fold increase in pancreatic protein output and a >10-fold increase in plasma CCK. Luminal DBI levels increased 4-fold. At 5 hours after diversion of bile-pancreatic juice, each of these measures returned to basal levels. Intraduodenal infusion of peptone evoked a 5-fold increase in the concentration of luminal DBI. In separate studies, it was demonstrated that intraduodenal administration of antiserum to a DBI peptide specifically abolished pancreatic secretion and the increase in plasma CCK levels after diversion of bile-pancreatic juice. To demonstrate that DBI mediates the postprandial rise in plasma CCK levels, it was shown that intraduodenal administration of 5% peptone induced dramatic increases in pancreatic secretion and plasma CCK, effects that could be blocked by intraduodenal administration of anti-DBI antiserum. Hence, DBI, a trypsin-sensitive CCK-RP secreted from the proximal small bowel, mediates the feedback regulation of pancreatic secretion and the postprandial release of CCK.

Diazepam-binding inhibitor-derived peptides induce intracellular calcium changes and modulate human neutrophil function. The effects of two diazepam-binding inhibitor (DBI)-derived peptides, triakontatetrapeptide (DBI 17-50, TTN) and eiksoneuropeptide (DBI 51-70, ENP), on cytosolic free Ca^{2+} concentrations ($[\text{Ca}^{2+}]_i$), chemotaxis, superoxide anion (O_2^-) generation, and phagocytosis in human neutrophils were studied. Both TTN and ENP induced a rapid and transient rise of $[\text{Ca}^{2+}]_i$. The effect of TTN depended on the presence of extracellular Ca^{2+} , whereas the effect of ENP also persisted after extracellular Ca^{2+} chelation. TTN induced neutrophil chemotaxis, stimulated O_2^- generation, and enhanced phagocytosis. ENP did not affect cell migration and oxidative metabolism but enhanced phagocytosis. Both peptides modulated N-formyl-methionyl-leucyl-phenylalanine- and phorbol myristate acetate-induced O_2^- generation. Because neutrophils express benzodiazepine receptors of the peripheral type (pBRs) and DBI-derived peptides may interact with such receptors, the possible role of pBRs in TTN- or ENP-induced effects were studied. The synthetic pBR ligand RO 5-4864 increased $[\text{Ca}^{2+}]_i$ through extracellular Ca^{2+} influx and this effect was prevented by the pBR antagonist PK-11195. RO 5-4864, however, was ineffective on neutrophil migration and O_2^- generation and only slightly affected phagocytosis. Moreover, PK-11195 delayed the $[\text{Ca}^{2+}]_i$ rise induced by TTN but did not significantly affect its extent, and had no effect on the $[\text{Ca}^{2+}]_i$ rise induced by ENP. DBI-derived peptides induce $[\text{Ca}^{2+}]_i$ changes and modulate neutrophil function mainly through pBR-independent pathways. In view of the wide cell and tissue distribution of DBI in the brain and in peripheral organs, modulation of

neutrophil function by DBI-derived peptides may be relevant for both the neuroimmune network and the development and regulation of the inflammatory processes.

Peripheral benzodiazepine (BDZ) receptor (PBR) and diazepam-binding inhibitor/acyl-CoA-binding protein (DBI/ACBP) characterized as a ligand at central BDZ receptors, at PBR with involvement in the regulation of steroidogenesis, and as an intracellular acyl-CoA transporter, are both known to interact with BDZ in adult systems. Researchers investigated their expression after prenatal exposure to BDZ. Diazepam (1.25 mg/kg per day s.c.) was administered to time-pregnant Long Evans rats from gestational day (GD) 14 to 20. Expression of mRNAs encoding for PBR and for DBI/ACBP was studied in the same animals with (33)P-labeled 60 mer oligonucleotides (oligos) by in situ hybridization at GD20, and with (32)P-labeled oligos by Northern blot in steroidogenic and immune organs at postnatal day (PN) 14 and in adult offspring. Prenatal diazepam increased DBI/ACBP mRNA expression in male fetal adrenal and in fetal and PN14 testis. Thymus exhibited increased DBI/ACBP mRNA in male fetuses and in adult female offspring, and reduced organ weight at PN14 in both sexes. In female spleen, an increase in DBI/ACBP mRNA and a decrease in PBR mRNA was seen at PN14. Apart from the finding in spleen, no drug-induced changes in PBR mRNA were observed. The effects of prenatal diazepam were superimposed on treatment-independent sex differences in DBI/ACBP mRNA and PBR mRNA expression. Our data indicate that expression of DBI/ACBP mRNA in steroidogenic and immune organs can be affected by exposure to BDZ during ontogeny, while PBR mRNA expression appears to be less sensitive. They further reveal marked sex differences in the developmental patterns of the two proteins during pre- and postpubertal ontogeny.

The mechanisms underlying the increase in diazepam binding inhibitor (DBI) and its mRNA expression induced by nicotine (0.1 μ M) exposure for 24 h were studied using mouse cerebral cortical neurons in primary culture. Nicotine-induced (0.1 μ M) increases in DBI mRNA expression were abolished by hexamethonium, a nicotinic acetylcholine (nACh) receptor antagonist. Agents that stabilize the neuronal membrane, including tetrodotoxin (TTX), procainamide (a Na⁺) channel inhibitor), and local anesthetics (dibucaine and lidocaine), dose-dependently inhibited the increased expression of DBI mRNA by nicotine. The nicotine-induced increase in DBI mRNA expression was inhibited by L-type voltage-dependent Ca²⁺ channel (VDCC) inhibitors such as verapamil, calmodulin antagonist (W-7), and Ca²⁺/calmodulin-dependent protein kinase II (CAM II kinase) inhibitor (KN-62), whereas P/Q- and N-type VDCC inhibitors showed no effects. In addition, nicotine exposure for 24 h induced [³H]nicotine binding to the particulate fractions of the neurons with an

increased B(max) value and no changes in K(d). Under these conditions, the 30 mM KCl- and nicotine-induced $45\text{Ca}(2+)$ influx into the nicotine-treated neurons was significantly higher than those into non-treated neurons. These results suggest that the nicotine-stimulated increase in DBI mRNA expression is mediated by CAM II kinase activation resulting from the increase in intracellular $\text{Ca}(2+)$ through L-type VDCCs subsequent to the neuronal membrane depolarization associated with nACh receptor activation.

Benzodiazepine receptors and DBI play a major role in regulating steroid production in both the adrenals and central nervous system, and may be involved in the activation of the hypothalamic-pituitary-adrenal axis in stress response. Preliminary findings regarding the presence of plasmatic benzodiazepine binding inhibitory activity (BBIA) in 14 psychiatric patients prompted us to investigate it further in larger samples of psychiatric patients ($n = 44$) and in healthy controls ($n = 14$). The results have shown that BBIA is present in all the subjects included with statistically significant differences between the patients and controls. The highest concentrations were found in the patients with no difference between anxious or depressed patients, and the lowest in the healthy controls. These findings indicate that BBIA might play a role in the pathophysiology of some manifestations of anxiety.

The neuropeptides diazepam binding inhibitor (DBI) and corticotropin-releasing hormone (CRH) elicit anxietylike symptoms when administered intracerebroventricularly to laboratory animals. Because of the similarities between the symptoms of certain anxiety states and the alcohol withdrawal syndrome, suggesting that increased secretion of either of these endogenous neuropeptides may, at least in part, be responsible for the symptoms of alcohol withdrawal. DBI and CRH concentrations were measured in cerebrospinal fluid (CSF) of 15 alcohol-dependent patients during acute withdrawal (Day 1) and again at 3 week's abstinence (Day 21). In addition, plasma concentrations of cortisol were measured to evaluate the relationship between pituitary-adrenal axis activation and CSF CRH concentrations. CSF CRH ($p < .04$), but not CSF DBI, was significantly higher on Day 1 than on Day 21. Although there was a significant decrease in plasma cortisol from Day 1 to Day 21 ($p < .001$), a significant correlation between CSF CRH and plasma cortisol concentrations was not observed at either time point. Neither CSF neuropeptide correlated with clinical measures of withdrawal severity. These tentative findings may implicate CRH, but not DBI, in the pathogenesis of alcohol withdrawal. Alternately, the central release of CRH and DBI may not be adequately reflected in lumbar CSF.

Because diazepam binding inhibitor (DBI) and its processing products coexist with gamma-aminobutyric acid (GABA) in several axon terminals, DBI immunoreactivity was

measured in the cerebrospinal fluid (CSF) of individuals suffering from various neuropsychiatric disorders, that are believe to be associated with abnormalities of GABAergic transmission. Increased amounts of DBI-like immunoreactivity were found in the CSF of patients suffering from severe depression with a severe anxiety component (Barbaccia, Costa, Ferrero, Guidotti, Roy, Sunderland, Pickar, Paul and Goodwin, 1986). Moreover, the amount of DBI and its processing products was found to be increased in the CSF of patients with hepatic encephalopathy (HE) (Rothstein, McKhann, Guarneri, Barbaccia, Guidotti and Costa, 1989; Guarneri, Berkovich, Guidotti and Costa, 1990). The clinical rating of HE correlated with the extent of the increase in DBI in CSF. Other lines of research suggest that DBI and DBI processing products may be important factors in behavioral adaptation to stress, acting via benzodiazepine (BZD) binding sites, located on mitochondria. DBI and its processing products, ODN and TTN, are present in high concentrations in the hypothalamus and in the amygdala, two areas of the brain that are important in regulating behavioral patterns associated with conflict situations, anxiety and stress. In CSF, the content of DBI changes in association with corticotropin releasing factor (CRF) (Roy, Pickar, Gold, Barbaccia, Guidotti, Costa and Linnoila, 1989). Finally DBI is preferentially concentrated in steroidogenic tissues and cells (adrenal cortical cells, Leydig cells of the testes and glial cells of the brain).

The disclosed NOV1 nucleic acid of the invention encoding a endozepine-related protein precursor-like protein includes the nucleic acid whose sequence is provided in Table 1A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 1A while still encoding a protein that maintains its endozepine-related protein precursor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 10% percent of the bases may be so changed.

The disclosed NOV1 protein of the invention includes the endozepine-related protein precursor-like protein whose sequence is provided in Table 1B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 1B while still encoding a protein that maintains its endozepine-related protein precursor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 60% percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this endozepine-related protein precursor-like protein (NOV1) may function as a member of a "endozepine-related protein precursor family". Therefore, the NOV1 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV1 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in cancer including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the endozepine-related protein precursor-like protein (NOV1) may be useful in gene therapy, and the endozepine-related protein precursor-like protein (NOV1) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, anxiety, depression, stress, immune dysfunction, alcoholism, obesity and diabetes. The NOV1 nucleic acid encoding the endozepine-related protein precursor-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV1 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV1 substances for use in therapeutic or

diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV1 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated NOV1 epitope is from about amino acids 480 to 500. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a “mature” form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product “mature” form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising

from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event.

5 Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as
10 approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization
15 technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA
20 of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can
25 be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence SEQ ID NO: 1, or a complement of this aforementioned nucleotide
30 sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55 as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A

LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

5 A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

10 As used herein, the term “oligonucleotide” refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue.

15 Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, or a complement thereof.

20 Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment
25 encoding a biologically-active portion of an NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55 is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55 that it can hydrogen bond with little or no
30 mismatches to the nucleotide sequence shown SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55, thereby forming a stable duplex.

As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated

polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA.

Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

An NOVX polypeptide is encoded by the open reading frame ("ORF") of an NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55; or an anti-sense strand nucleotide sequence of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55; or of a naturally occurring mutant of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group

can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express an NOVX protein, such as by measuring a level of an NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of an NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, that encodes a polypeptide having an NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding an NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the

NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of 1. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in

which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at

pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or

oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and

5 oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y.

(1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%,

10 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain

hybridized to each other. A non-limiting example of stringent hybridization conditions are

hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated

15 nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

20 In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's

25 solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

30 In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM

EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See*,
5 *e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

10 **Conservative Mutations**

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, thereby leading to changes in the amino acid sequences of
15 the encoded NOVX proteins, without altering the functional ability of said NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their
20 biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX
25 proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous
30 to the amino acid sequences SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56; more preferably at least about 70% homologous SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56; still more preferably at least about

80% homologous to SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56; even more preferably at least about 90% homologous to SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56; and most preferably at least about 95% homologous to SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56.

An isolated nucleic acid molecule encoding an NOVX protein homologous to the protein of SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis, SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any

one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein and an NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (e.g., regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of an NOVX protein of SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56, or antisense nucleic acids complementary to an NOVX nucleic acid sequence of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which

flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or

genomic DNA encoding an NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330).

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having

specificity for an NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of an NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an NOVX-encoding mRNA. *See*, *e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S₁ nucleases (*See*, Hyrup, *et al.*, 1996. *supra*); or as probes or primers for DNA sequence and hybridization (*See*, Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the

formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see*, Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52,

54, or 56, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, an NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or

other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of an NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of an NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56, and retains the functional activity of the protein of SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56, and retains the functional activity of the NOVX proteins of SEQ ID NO: 2, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a

second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, an NOVX "chimeric protein" or "fusion protein" comprises an NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an NOVX protein SEQ ID NO: 2, 22, 24, 26, 28, 30,

32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, or 56, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within an NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of an NOVX protein. In one embodiment, an NOVX fusion protein comprises at least one biologically-active portion of an NOVX protein. In another embodiment, an NOVX fusion protein comprises at least two biologically-active portions of an NOVX protein. In yet another embodiment, an NOVX fusion protein comprises at least three biologically-active portions of an NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is an NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an NOVX ligand and an NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of an NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with an NOVX ligand.

An NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene

library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of an NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates

isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. See, e.g., Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, et al., 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, F_{ab'} and F_{(ab')₂} fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX-related protein that is located on the surface of the protein, e.g., a hydrophilic region. A hydrophobicity analysis of the human NOVX-related

protein sequence will indicate which regions of a NOVX-related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, e.g., Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow and Lane, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and *Corynebacterium parvum*, or similar immunostimulatory agents.

Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

5 The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson
10 (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

15 The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MABs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the
20 antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically
25 bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell
30 line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that

preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur et al., MONOCLONAL ANTIBODY PRODUCTION TECHNIQUES AND APPLICATIONS, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by

using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, *Nature* 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., *Nature*, 321:522-525 (1986); Riechmann et al., *Nature*, 332:323-327 (1988); Verhoeven et al., *Science*, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that

of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)).

Human Antibodies

5 Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 *Immunol Today* 4: 72) and the EBV hybridoma
10 technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In:
15 MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the
20 endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992));
25 Lonberg et al. (*Nature* 368 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild et al, (*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's
30 endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human

DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and
5 WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human
10 variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent
15 No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and
20 germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into
25 another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT
30 publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent

No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the
5 idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F_{(ab')₂} fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_{(ab')₂} fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is
15 any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random
20 assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

25 Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the
30 fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced

at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments
5 comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv
10 (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm
15 of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc γ R), such as Fc γ RI (CD64), Fc γ RII (CD32) and Fc γ RIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular
20 antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

25 **Heteroconjugate Antibodies**

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO
30 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate

and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

5 It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and
10 antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis
15 and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active
20 toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain
25 (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples
30 include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as

glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in
5 Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is
10 administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and
15 other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided
20 herein.

Anti-NOVX antibodies may be used in methods known within the art relating to the localization and/or quantitation of an NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for NOVX
25 proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-NOVX antibody (*e.g.*, monoclonal antibody) can be used to isolate an NOVX polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation.
30 An anti-NOVX antibody can facilitate the purification of natural NOVX polypeptide from cells and of recombinantly-produced NOVX polypeptide expressed in host cells. Moreover, an anti-NOVX antibody can be used to detect NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the NOVX protein. Anti-NOVX antibodies can be used diagnostically to monitor protein levels in tissue

as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that

the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the

recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL
5 (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION
10 TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.*, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY
15 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.*, Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

20 In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

25 Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian
30 cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable

expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

5 In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

20 The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see, *e.g.*, Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and

"recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable

medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals

via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55 can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX

gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g., Li, et al., 1992. Cell 69: 915.*

The selected cells are then injected into a blastocyst of an animal (*e.g., a mouse*) to form aggregation chimeras. *See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152.* A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol. 2: 823-829*; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g., Lakso, et al., 1992. Proc. Natl. Acad. Sci. USA 89: 6232-6236.* Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See, O'Gorman, et al., 1991. Science 251:1351-1355.* If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g., by* mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al., 1997. Nature 385: 810-813.* In brief, a cell (*e.g., a somatic cell*) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, *e.g., through the use of* electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g., the somatic cell*) is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF,

Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, an NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a

lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, serotonin receptor, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be

achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in an NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease (possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of an NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science*

249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to an NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule. As used herein, a "target molecule" is a molecule with which an NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An NOVX target molecule can be a non-NOVX molecule or an

NOVX protein or polypeptide of the invention. In one embodiment, an NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be
5 a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or
10 interact with an NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+} , diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising an NOVX-responsive
15 regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting an NOVX protein or biologically-active portion thereof with a test compound and
20 determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test
25 compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX
30 protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to an NOVX target molecule by one of the methods described

above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate an NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of an NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)_n, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is

incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8:

5 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors,
10 which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation
15 domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be
20 detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

25 Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic
30 identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences, SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g., D'Eustachio, et al., 1983. Science 220: 919-924.* Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one

step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually.

- 5 The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see*, Verma, *et al.*, HUMAN CHROMOSOMES: A
10 MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more
15 likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line
20 through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and
25 unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome
30 spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in an NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological

sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO: 1, 21, 23, 25, 27, 28, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, or 55, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test

subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be

used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (e.g., wherein
5 the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in an NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder
10 characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding an NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of
15 one or more nucleotides from an NOVX gene; (ii) an addition of one or more nucleotides to an NOVX gene; (iii) a substitution of one or more nucleotides of an NOVX gene, (iv) a chromosomal rearrangement of an NOVX gene; (v) an alteration in the level of a messenger RNA transcript of an NOVX gene, (vi) aberrant modification of an NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern
20 of a messenger RNA transcript of an NOVX gene, (viii) a non-wild-type level of an NOVX protein, (ix) allelic loss of an NOVX gene, and (x) inappropriate post-translational modification of an NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in an NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional
25 means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g.,
30 Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (see, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the

nucleic acid sample with one or more primers that specifically hybridize to an NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see*, Kwok, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q β Replicase (*see*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.*, Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes.

This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence.

Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (see, e.g., Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S₁ nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. See, e.g., Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on an NOVX sequence, e.g., a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is

treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79.

Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.,* NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such treatment, the pharmacogenomics (*i.e.,* the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may

be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid

metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified.

Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods

as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

5 In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression
10 of an NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein,
15 mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to
20 lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include cardiomyopathy, atherosclerosis,
25 hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation,
30 idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, and other diseases, disorders and conditions of the like.

These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with

- 5 Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid
- 10 and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies
- 15 specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity

20 may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

- Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for
- 25 RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in*
- 30 *situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the

subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, an NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of an NOVX protein, a peptide, an NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering an NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect.

One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (e.g., cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (e.g., preclampsia).

Determination of the Biological Effect of the Therapeutic

5 In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

10 In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

15 Prophylactic and Therapeutic Uses of the Compositions of the Invention

20 The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

25 As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

30 Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial

106

properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit
5 the scope of the invention described in the claims.

Examples

Example 1. Identification of NOVX clones

The novel NOVX target sequences identified in the present invention were subjected to
10 the exon linking process to confirm the sequence. PCR primers were designed by starting at
the most upstream sequence available, for the forward primer, and at the most downstream
sequence available for the reverse primer. Table 2A shows the sequences of the PCR primers
used for obtaining different clones. In each case, the sequence was examined, walking inward
from the respective termini toward the coding sequence, until a suitable sequence that is either
15 unique or highly selective was encountered, or, in the case of the reverse primer, until the stop
codon was reached. Such primers were designed based on in silico predictions for the full
length cDNA, part (one or more exons) of the DNA or protein sequence of the target
sequence, or by translated homology of the predicted exons to closely related human
sequences from other species. These primers were then employed in PCR amplification based
20 on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain
- cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole,
fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary
gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small
intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting
25 amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product
derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting
bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1
vector. Table 2B shows a list of these bacterial clones. The resulting sequences from all
clones were assembled with themselves, with other fragments in CuraGen Corporation's
30 database and with public ESTs. Fragments and ESTs were included as components for an
assembly when the extent of their identity with another component of the assembly was at
least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for
corrections if appropriate. These procedures provide the sequence reported herein.

Table 2A. PCR Primers for Exon Linking

NOVX Clone	Primer 1 (5' - 3')	SEQ ID NO	Primer 2 (5' - 3')	SEQ ID NO
NOV1	CTCCTTTTGGGGCATGTTGATCC	9	GATTTTCTTGTGAACACCACAATCCAG	10

Physical clone: Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

Table 2B. Physical Clones for PCR products

NOVX Clone	Bacterial Clone
NOV2	58303::20936375..698050.K10 FLC ELT

Example 2. Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), AI_comprehensive_panel (containing normal tissue and samples from autoinflammatory diseases), Panel CNSD.01 (containing samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and diseased brains).

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β -actin and GAPDH). Normalized RNA (5 μ l) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before

selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60° C, primer optimal T_m = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe T_m must be 10° C greater than primer T_m , amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthesgen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl₂, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/μl RNase inhibitor, and 0.25 U/μl reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

Panel 1

In the results for Panel 1, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

Panel 2

5 The plates for Panel 2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast

cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA
5 extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be
10 indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 4

15 Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene ,La Jolla, CA) and thymus and kidney (Clontech) were employed.
20 Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells,
25 small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as
30 indicated. The following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2 μ g/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at 5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 μ g/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2×10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol (5.5×10^{-5} M) (Gibco), and 10 mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions.

Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml GM-CSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10 mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100 ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 μ g/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. Then

CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and plated at 10^6 cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 μ g/ml anti-CD28 (Pharmingen) and 3 μ g/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 μ g/ml or anti-CD40 (Pharmingen) at approximately 10 μ g/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at 24, 48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μ g/ml anti-CD28 (Pharmingen) and 2 μ g/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5 - 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1 μ g/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1 μ g/ml) were used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium

pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 μ g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at 5×10^5 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5×10^5 cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10 mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1 μ g/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at -20 degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 μ l of RNase-free water and 35 μ l buffer (Promega) 5 μ l DTT, 7 μ l RNasin and 8 μ l DNase were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume

of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at -80 degrees C.

Panel CNSD.01

5

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor.

10 All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

30 In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

35 Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

5

The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

10

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented:

15

Hippocampus, Temporal cortex (Brodmann Area 21), Somatosensory cortex (Brodmann area 7), and Occipital cortex (Brodmann area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the somatosensory cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

20

25

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

30

In the labels employed to identify tissues in the CNS_Neurodegeneration_V1.0 panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

Control (Path) = Control brains; patient not demented but showing severe AD-like pathology

SupTemporal Ctx = Superior Temporal Cortex

Inf Temporal Ctx = Inferior Temporal Cortex

NOV1

Expression of the NOV1 gene (20936375_0_228_da1) was assessed using the primer-probe sets Ag1865, Ag2029, and Ag2813 described in Tables 3, 4 and 5. Results from RTQ-PCR runs are shown in Tables 6, 7, 8, 9 and 10.

Table 3. Probe Name Ag1377

Primers	Sequences	TM	Length	Start Position
Forward	5'-TATTACTTGGGTGGTCATTCCA-3' (SEQ ID NO: 11)	59.2	22	1003
Probe	FAM-5'-CAACCCATGGAATTCCTGGATTTCG-3'-TAMRA (SEQ ID NO: 12)	68.9	26	1027
Reverse	5'-ATATTCCTCAATGTTGCCATTTC-3' (SEQ ID NO: 13)	59.9	22	1076

Table 4. Probe Name Ag1865/Ag2029 (identical sequences)

Primers	Sequences	TM	Length	Start Position
Forward	5'-AGGCAAAATCATCAACATCAAC-3' (SEQ ID NO: 14)	59	22	1434
Probe	TET-5'-CTCAGCCACCTCACAGAGACCATCT-3'-TAMRA (SEQ ID NO: 15)	70	26	1470
Reverse	5'-TTAGCACACCAGGAGACATCTC-3' (SEQ ID NO: 16)	59.4	22	1508

Table 5. Probe Name Ag2813

Primers	Sequences	TM	Length	Start Position
Forward	5'-AATCATCAACATCAACATTGCA-3' (SEQ ID NO: 17)	58.9	22	1564
Probe	FAM-5'-CTCAGCCACCTCACAGAGACCATCT-3'-TAMRA (SEQ ID NO: 18)	70	26	1594
Reverse	5'-GTTAGCACACCAGGAGACATCT-3' (SEQ ID NO: 19)	58.3	22	1633

Table 6. Panel 1.2

Tissue Name	Relative Expression(%)	Tissue Name	Relative Expression(%)
	1.2tm1607f_ ag1377		1.2tm1607f_ ag1377
Endothelial cells	3.7	Renal ca. 786-0	32.3
Heart (fetal)	2.8	Renal ca. A498	8.7
Pancreas	2.6	Renal ca. RXF 393	1.6
Pancreatic ca. CAPAN 2	4.7	Renal ca. ACHN	3.1
Adrenal Gland (new lot*)	16.7	Renal ca. UO-31	1.3
Thyroid	8.0	Renal ca. TK-10	2.4
Salivary gland	23.5	Liver	21.8
Pituitary gland	3.9	Liver (fetal)	5.3
Brain (fetal)	7.1	Liver ca. (hepatoblast) HepG2	12.2
Brain (whole)	19.3	Lung	5.0
Brain (amygdala)	12.3	Lung (fetal)	4.7
Brain (cerebellum)	10.1	Lung ca. (small cell) LX-1	16.7
Brain (hippocampus)	24.7	Lung ca. (small cell) NCI-H69	6.2
Brain (thalamus)	15.0	Lung ca. (s.cell var.) SHP-77	1.9
Cerebral Cortex	24.5	Lung ca. (large cell)NCI-H460	16.0
Spinal cord	7.4	Lung ca. (non-sm. cell) A549	9.3
CNS ca. (glio/astro) U87-MG	12.7	Lung ca. (non-s.cell) NCI-H23	1.9
CNS ca. (glio/astro) U-118-MG	2.2	Lung ca (non-s.cell) HOP-62	2.9
CNS ca. (astro) SW1783	3.5	Lung ca. (non-s.cl) NCI-H522	11.8
CNS ca.* (neuro; met) SK-N-AS	100.0	Lung ca. (squam.) SW 900	8.0
CNS ca. (astro) SF-539	2.9	Lung ca. (squam.) NCI-H596	1.6
CNS ca. (astro) SNB-75	0.8	Mammary gland	4.4
CNS ca. (glio) SNB-19	5.7	Breast ca.* (pl. effusion) MCF-7	39.8
CNS ca. (glio) U251	0.7	Breast ca.* (pl.ef) MDA-MB-231	14.6
CNS ca. (glio) SF-295	0.9	Breast ca.* (pl. effusion) T47D	22.4
Heart	48.3	Breast ca. BT-549	15.8
Skeletal Muscle (new lot*)	43.2	Breast ca. MDA-N	23.0
Bone marrow	3.3	Ovary	0.9
Thymus	2.6	Ovarian ca. OVCAR-3	4.3
Spleen	4.9	Ovarian ca. OVCAR-4	4.7
Lymph node	3.7	Ovarian ca. OVCAR-5	13.4
Colorectal	1.6	Ovarian ca. OVCAR-8	20.0
Stomach	6.3	Ovarian ca. IGROV-1	6.9
Small intestine	22.2	Ovarian ca.* (ascites) SK-OV-3	5.6
Colon ca. SW480	0.7	Uterus	6.9
Colon ca.* (SW480 met)SW620	8.5	Placenta	30.1
Colon ca. HT29	14.4	Prostate	10.3
Colon ca. HCT-116	6.6	Prostate ca.* (bone met)PC-3	14.3
Colon ca. CaCo-2	21.6	Testis	4.1
83219 CC Well to Mod Diff	3.2	Melanoma Hs688(A).T	1.5

(ODO3866)			
Colon ca. HCC-2998	42.9	Melanoma* (met) Hs688(B).T	1.0
Gastric ca.* (liver met) NCI-N87	31.2	Melanoma UACC-62	1.4
Bladder	9.3	Melanoma M14	2.1
Trachea	4.0	Melanoma LOX IMVI	0.8
Kidney	25.2	Melanoma* (met) SK-MEL-5	5.6
Kidney (fetal)	9.3		

Table 7. Panel 1.3D

Tissue Name	Relative Expression(%)	Relative Expression(%)	Relative Expression(%)
	1.3dtm2944t_a g1865	1.3dx4tm5428t ag2029_b1	1.3dx4tm5380f ag2813_b2
Liver adenocarcinoma	10.2	4.3	9.1
Pancreas	5.0	8.3	6.9
Pancreatic ca. CAPAN 2	8.9	20.3	24.7
Adrenal gland	13.2	17.8	17.8
Thyroid	9.7	15.0	11.2
Salivary gland	3.9	8.2	11.4
Pituitary gland	7.6	9.6	8.7
Brain (fetal)	10.2	34.8	36.2
Brain (whole)	43.5	100.0	100.0
Brain (amygdala)	36.1	61.3	44.7
Brain (cerebellum)	10.0	54.5	38.1
Brain (hippocampus)	100.0	57.4	63.6
Brain (substantia nigra)	12.4	43.1	45.4
Brain (thalamus)	34.4	72.1	81.9
Cerebral Cortex	50.0	13.6	11.6
Spinal cord	19.6	68.8	44.1
CNS ca. (glio/astro) U87-MG	6.7	13.5	10.7
CNS ca. (glio/astro) U-118-MG	13.5	32.7	19.8
CNS ca. (astro) SW1783	8.7	13.2	11.1
CNS ca.* (neuro; met) SK-N-AS	52.9	56.5	51.5
CNS ca. (astro) SF-539	9.6	11.9	14.4
CNS ca. (astro) SNB-75	21.8	22.9	20.2
CNS ca. (glio) SNB-19	11.2	17.1	22.6
CNS ca. (glio) U251	3.2	19.8	14.3
CNS ca. (glio) SF-295	10.3	6.5	8.7
Heart (fetal)	6.4	1.0	0.7
Heart	7.5	30.4	34.2
Fetal Skeletal	19.9	0.0	0.4
Skeletal muscle	6.7	73.6	52.3
Bone marrow	4.0	6.3	5.4
Thymus	4.5	5.1	6.4
Spleen	14.7	12.6	11.6

Lymph node	9.2	23.3	21.7
Colorectal	8.4	3.2	3.9
Stomach	20.0	17.3	11.4
Small intestine	12.8	31.4	18.9
Colon ca. SW480	10.0	3.9	4.3
Colon ca.* (SW480 met)SW620	8.3	7.7	9.7
Colon ca. HT29	12.3	6.6	9.6
Colon ca. HCT-116	6.8	6.6	8.9
Colon ca. CaCo-2	25.3	16.4	17.4
83219 CC Well to Mod Diff (ODO3866)	8.1	16.1	14.5
Colon ca. HCC-2998	20.9	19.0	19.5
Gastric ca.* (liver met) NCI-N87	20.0	31.2	30.6
Bladder	11.3	17.7	11.8
Trachea	13.2	10.9	10.9
Kidney	7.9	23.8	18.2
Kidney (fetal)	6.7	13.5	14.7
Renal ca. 786-0	24.0	30.4	28.8
Renal ca. A498	24.5	37.8	36.8
Renal ca. RXF 393	5.2	15.4	19.6
Renal ca. ACHN	10.6	13.7	12.6
Renal ca. UO-31	12.3	20.2	16.8
Renal ca. TK-10	3.2	12.1	8.7
Liver	10.7	23.0	30.0
Liver (fetal)	6.9	19.3	16.3
Liver ca. (hepatoblast) HepG2	13.0	34.1	32.8
Lung	9.0	13.0	11.9
Lung (fetal)	12.6	13.6	19.6
Lung ca. (small cell) LX-1	14.6	30.0	29.6
Lung ca. (small cell) NCI-H69	10.4	6.0	7.8
Lung ca. (s.cell var.) SHP-77	11.5	23.9	15.9
Lung ca. (large cell)NCI-H460	5.2	25.3	21.7
Lung ca. (non-sm. cell) A549	12.9	8.0	14.6
Lung ca. (non-s.cell) NCI-H23	14.9	8.9	5.3
Lung ca. (non-s.cell) HOP-62	4.3	5.1	8.2
Lung ca. (non-s.cl) NCI-H522	24.1	11.0	14.8
Lung ca. (squam.) SW 900	9.0	19.9	20.7
Lung ca. (squam.) NCI-H596	3.0	6.3	6.9
Mammary gland	11.2	14.2	9.6
Breast ca.* (pl. effusion) MCF-7	13.5	34.7	21.9
Breast ca.* (pl.ef) MDA-MB-231	15.0	46.8	37.9
Breast ca.* (pl. effusion) T47D	7.0	8.7	6.7
Breast ca. BT-549	18.0	34.9	28.3
Breast ca. MDA-N	13.1	8.3	11.4
Ovary	9.0	0.6	0.8
Ovarian ca. OVCAR-3	6.3	17.3	20.8

Ovarian ca. OVCAR-4	2.8	17.6	13.7
Ovarian ca. OVCAR-5	20.3	23.2	21.2
Ovarian ca. OVCAR-8	11.7	7.4	9.0
Ovarian ca. IGROV-1	7.6	6.7	7.8
Ovarian ca.* (ascites) SK-OV-3	15.8	24.1	37.4
Uterus	5.6	20.9	17.4
Placenta	6.2	14.9	8.4
Prostate	7.4	13.3	5.8
Prostate ca.* (bone met)PC-3	8.2	7.8	9.9
Testis	58.2	19.0	36.3
Melanoma Hs688(A).T	7.4	4.7	4.2
Melanoma* (met) Hs688(B).T	8.1	4.1	4.0
Melanoma UACC-62	0.4	2.5	2.9
Melanoma M14	9.5	32.4	34.5
Melanoma LOX IMVI	2.3	1.4	2.0
Melanoma* (met) SK-MEL-5	7.1	8.8	12.1
Adipose	8.6	9.3	8.0

Table 8. Panel 2D

Tissue Name	Relative Expression(%)	Tissue Name	Relative Expression(%)
	2dx4tm4710f_ag2813_a1		2dx4tm4710f_ag2813_a1
Normal Colon GENPAK 061003	56.8	Kidney NAT Clontech 8120608	4.5
83219 CC Well to Mod Diff (ODO3866)	18.8	Kidney Cancer Clontech 8120613	11.2
83220 CC NAT (ODO3866)	14.3	Kidney NAT Clontech 8120614	6.8
83221 CC Gr.2 rectosigmoid (ODO3868)	13.4	Kidney Cancer Clontech 9010320	6.9
83222 CC NAT (ODO3868)	3.4	Kidney NAT Clontech 9010321	11.9
83235 CC Mod Diff (ODO3920)	10.9	Normal Uterus GENPAK 061018	3.5
83236 CC NAT (ODO3920)	9.4	Uterus Cancer GENPAK 064011	18.8
83237 CC Gr.2 ascend colon (ODO3921)	28.2	Normal Thyroid Clontech A+ 6570-1	21.4
83238 CC NAT (ODO3921)	9.8	Thyroid Cancer GENPAK 064010	16.9
83241 CC from Partial Hepatectomy (ODO4309)	60.7	Thyroid Cancer INVITROGEN A302152	20.3
83242 Liver NAT (ODO4309)	56.5	Thyroid NAT INVITROGEN A302153	21.7
87472 Colon mets to lung (OD04451-01)	8.7	Normal Breast GENPAK 061019	16.6
87473 Lung NAT (OD04451-02)	8.1	84877 Breast Cancer (OD04566)	13.2
Normal Prostate Clontech A+ 6546-1	100.0	85975 Breast Cancer (OD04590-01)	50.9

<u>84140 Prostate Cancer (OD04410)</u>	31.7	<u>85976 Breast Cancer Mets (OD04590-03)</u>	54.2
<u>84141 Prostate NAT (OD04410)</u>	27.0	<u>87070 Breast Cancer Metastasis (OD04655-05)</u>	37.9
<u>87073 Prostate Cancer (OD04720-01)</u>	10.0	GENPAK Breast Cancer 064006	11.3
<u>87074 Prostate NAT (OD04720-02)</u>	19.5	Breast Cancer Res. Gen. 1024	13.7
Normal Lung GENPAK 061010	35.6	Breast Cancer Clontech 9100266	24.0
<u>83239 Lung Met to Muscle (ODO4286)</u>	20.4	Breast NAT Clontech 9100265	6.8
<u>83240 Muscle NAT (ODO4286)</u>	8.6	Breast Cancer INVITROGEN A209073	17.0
<u>84136 Lung Malignant Cancer (OD03126)</u>	29.7	Breast NAT INVITROGEN A2090734	8.3
<u>84137 Lung NAT (OD03126)</u>	26.6	Normal Liver GENPAK 061009	36.7
<u>84871 Lung Cancer (OD04404)</u>	17.9	Liver Cancer GENPAK 064003	19.2
<u>84872 Lung NAT (OD04404)</u>	10.8	Liver Cancer Research Genetics RNA 1025	15.8
<u>84875 Lung Cancer (OD04565)</u>	4.7	Liver Cancer Research Genetics RNA 1026	4.8
<u>84876 Lung NAT (OD04565)</u>	5.1	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	18.0
<u>85950 Lung Cancer (OD04237-01)</u>	32.1	Paired Liver Tissue Research Genetics RNA 6004-N	9.1
<u>85970 Lung NAT (OD04237-02)</u>	19.5	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	7.6
<u>83255 Ocular Mel Met to Liver (ODO4310)</u>	9.8	Paired Liver Tissue Research Genetics RNA 6005-N	4.9
<u>83256 Liver NAT (ODO4310)</u>	44.9	Normal Bladder GENPAK 061001	25.8
<u>84139 Melanoma Mets to Lung (OD04321)</u>	9.4	Bladder Cancer Research Genetics RNA 1023	4.6
<u>84138 Lung NAT (OD04321)</u>	18.6	Bladder Cancer INVITROGEN A302173	11.1
Normal Kidney GENPAK 061008	69.2	<u>87071 Bladder Cancer (OD04718-01)</u>	31.5
<u>83786 Kidney Ca, Nuclear grade 2 (OD04338)</u>	58.5	<u>87072 Bladder Normal Adjacent (OD04718-03)</u>	9.6
<u>83787 Kidney NAT (OD04338)</u>	24.6	Normal Ovary Res. Gen.	1.7
<u>83788 Kidney Ca Nuclear grade 1/2 (OD04339)</u>	17.3	Ovarian Cancer GENPAK 064008	18.7
<u>83789 Kidney NAT (OD04339)</u>	58.1	<u>87492 Ovary Cancer (OD04768-07)</u>	65.7
<u>83790 Kidney Ca, Clear cell type (OD04340)</u>	36.5	<u>87493 Ovary NAT (OD04768-08)</u>	4.4
<u>83791 Kidney NAT (OD04340)</u>	33.9	Normal Stomach GENPAK 061017	12.9
<u>83792 Kidney Ca, Nuclear</u>	9.8	Gastric Cancer Clontech	3.4

grade 3 (OD04348)		9060358	
83793 Kidney NAT (OD04348)	29.2	NAT Stomach Clontech 9060359	12.3
87474 Kidney Cancer (OD04622-01)	19.1	Gastric Cancer Clontech 9060395	13.7
87475 Kidney NAT (OD04622- 03)	4.6	NAT Stomach Clontech 9060394	9.8
85973 Kidney Cancer (OD04450-01)	49.5	Gastric Cancer Clontech 9060397	36.0
85974 Kidney NAT (OD04450- 03)	44.4	NAT Stomach Clontech 9060396	2.6
Kidney Cancer Clontech 8120607	3.1	Gastric Cancer GENPAK 064005	70.8

Table 9. Panel 4D

Tissue Name	Relative Expression(%)	Relative Expression(%)	Relative Expression(%)	Relative Expression(%)
	4dtm2444f_ag 1377	4Dx4tm4532t ag1865_a1	4dx4tm4450t ag2029_a2	4dx4tm4579f ag2813_a2
93768_Secondary Th1_anti- CD28/anti-CD3	20.2	22.3	15.9	17.2
93769_Secondary Th2_anti- CD28/anti-CD3	14.7	25.7	17.6	26.7
93770_Secondary Tr1_anti- CD28/anti-CD3	21.2	15.5	20.8	18.9
93573_Secondary Th1_resting day 4-6 in IL-2	7.9	10.1	4.5	8.4
93572_Secondary Th2_resting day 4-6 in IL-2	12.6	8.6	9.3	10.8
93571_Secondary Tr1_resting day 4-6 in IL-2	8.0	14.7	9.8	13.8
93568_primary Th1_anti- CD28/anti-CD3	22.8	20.3	16.0	18.4
93569_primary Th2_anti- CD28/anti-CD3	23.3	28.0	15.6	18.4
93570_primary Tr1_anti- CD28/anti-CD3	30.8	24.1	17.2	26.3
93565_primary Th1_resting dy 4-6 in IL-2	46.7	48.9	33.7	48.9
93566_primary Th2_resting dy 4-6 in IL-2	19.2	17.7	18.5	23.8
93567_primary Tr1_resting dy 4-6 in IL-2	14.1	18.2	13.8	22.2
93351_CD45RA CD4 lymphocyte_anti-CD28/anti- CD3	12.1	17.8	9.3	13.4
93352_CD45RO CD4 lymphocyte_anti-CD28/anti- CD3	20.7	33.9	14.3	24.0
93251_CD8 Lymphocytes_anti- CD28/anti-CD3	16.6	17.2	13.3	18.2
93353_chronic CD8	18.3	22.3	16.1	20.0

Lymphocytes 2ry_resting dy 4-6 in IL-2				
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	14.4	17.4	12.1	13.1
93354_CD4_none	6.5	9.2	6.8	5.7
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	19.5	17.0	13.0	20.1
93103_LAK cells_resting	35.1	33.7	24.7	37.0
93788_LAK cells_IL-2	30.6	34.7	22.8	30.5
93787_LAK cells_IL-2+IL-12	21.3	21.3	15.1	17.8
93789_LAK cells_IL-2+IFN gamma	31.4	36.8	23.7	34.3
93790_LAK cells_IL-2+ IL-18	20.9	21.7	17.9	26.3
93104_LAK cells_PMA/ionomycin and IL-18	9.4	13.5	11.0	13.8
93578_NK Cells IL-2_resting	11.7	32.4	17.7	28.9
93109_Mixed Lymphocyte Reaction_Two Way MLR	19.9	56.4	21.9	26.1
93110_Mixed Lymphocyte Reaction_Two Way MLR	12.4	18.7	14.3	17.9
93111_Mixed Lymphocyte Reaction_Two Way MLR	8.6	13.4	6.9	10.5
93112_Mononuclear Cells (PBMcs)_resting	9.9	18.0	9.6	12.9
93113_Mononuclear Cells (PBMcs)_PWM	70.7	86.0	53.7	74.4
93114_Mononuclear Cells (PBMcs)_PHA-L	35.4	31.9	14.4	24.4
93249_Ramos (B cell)_none	27.5	21.6	16.8	23.3
93250_Ramos (B cell)_ionomycin	72.2	97.4	80.9	100.0
93349_B lymphocytes_PWM	57.8	100.0	100.0	98.0
93350_B lymphocytes_CD40L and IL-4	15.0	23.4	20.0	28.9
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	7.0	18.2	12.1	15.1
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	9.0	19.1	14.9	18.1
93356_Dendritic Cells_none	19.6	35.9	21.3	29.6
93355_Dendritic Cells_LPS 100 ng/ml	15.5	36.3	36.4	43.9
93775_Dendritic Cells_anti-CD40	13.9	39.4	24.3	21.5
93774_Monocytes_resting	21.2	28.7	19.9	20.3
93776_Monocytes_LPS 50 ng/ml	22.1	30.2	23.9	21.9
93581_Macrophages_resting	26.4	38.1	30.7	34.3

93582_Macrophages_LPS 100 ng/ml	61.1	44.3	29.5	31.3
93098_HUVEC (Endothelial)_none	11.2	17.9	17.5	19.9
93099_HUVEC (Endothelial)_starved	26.2	38.8	19.8	34.5
93100_HUVEC (Endothelial)_IL-1b	4.9	12.9	7.3	7.1
93779_HUVEC (Endothelial)_IFN gamma	14.2	27.8	17.5	24.4
93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	6.5	28.5	21.0	26.9
93101_HUVEC (Endothelial)_TNF alpha + IL4	8.8	19.3	14.7	14.8
93781_HUVEC (Endothelial)_IL-11	7.1	11.1	7.4	7.6
93583_Lung Microvascular Endothelial Cells_none	13.6	32.1	21.4	26.7
93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	11.0	28.2	20.8	24.8
92662_Microvascular Dermal endothelium_none	34.4	35.3	31.6	38.7
92663_Microvascular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	15.7	25.0	18.9	23.5
93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	11.4	20.0	2.0	18.7
93347_Small Airway Epithelium_none	3.6	6.8	6.1	6.9
93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	26.6	40.2	30.2	35.2
92668_Coronary Artery SMC_resting	13.1	17.6	13.8	21.1
92669_Coronary Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	5.6	13.4	10.7	10.0
93107_astrocytes_resting	4.7	18.6	13.7	16.7
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	8.0	15.0	10.8	11.8
92666_KU-812 (Basophil)_resting	6.8	9.9	7.4	9.3
92667_KU-812 (Basophil)_PMA/ionoycin	20.6	31.3	20.9	26.0
93579_CCD1106 (Keratinocytes)_none	6.8	17.3	8.8	16.9
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	3.6	7.8	1.2	7.0
93791_Liver Cirrhosis	8.7	8.0	4.8	8.0

93792_Lupus Kidney	15.1	4.8	3.9	5.0
93577_NCI-H292	27.2	39.1	37.4	40.3
93358_NCI-H292_IL-4	34.2	54.0	43.5	53.9
93360_NCI-H292_IL-9	30.8	52.9	47.9	56.9
93359_NCI-H292_IL-13	18.0	40.7	29.0	16.2
93357_NCI-H292_IFN gamma	13.6	49.2	39.6	47.8
93777_HPAEC_-	10.9	35.9	11.3	10.2
93778_HPAEC_IL-1 beta/TNA alpha	10.7	20.0	11.4	16.7
93254_Normal Human Lung Fibroblast_none	14.6	14.6	12.1	12.2
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	9.1	21.1	12.6	15.8
93257_Normal Human Lung Fibroblast_IL-4	11.6	24.0	19.9	23.5
93256_Normal Human Lung Fibroblast_IL-9	9.9	24.7	19.4	17.9
93255_Normal Human Lung Fibroblast_IL-13	21.5	14.1	9.7	12.8
93258_Normal Human Lung Fibroblast_IFN gamma	22.8	42.7	39.0	37.7
93106_Dermal Fibroblasts CCD1070_resting	20.9	37.6	31.6	34.0
93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	39.5	70.1	60.3	65.1
93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	11.2	26.2	18.8	18.6
93772_dermal fibroblast_IFN gamma	9.5	23.4	19.7	18.5
93771_dermal fibroblast_IL-4	14.5	24.9	16.4	22.9
93260_IBD Colitis 2	3.3	1.6	0.5	1.8
93261_IBD Crohns	3.4	4.5	2.7	4.2
735010_Colon_normal	26.6	52.5	41.6	53.0
735019_Lung_none	10.9	24.1	17.0	23.4
64028-1_Thymus_none	100.0	80.8	57.6	76.2
64030-1_Kidney_none	19.9	28.7	25.6	27.6

Table 10. CNS_neurodegeneration_panel_v1.0

Tissue Name	Relative Expression(%) tm7004t_ag18 65_a2_s1	Relative Expression(%) tm7053f_ag28 13_b1_s2
AD 1 Hippo	13.4	14.0
AD 2 Hippo	30.9	33.3
AD 3 Hippo	9.0	12.1
AD 4 Hippo	12.3	13.9
AD 5 Hippo	89.8	100.0
AD 6 Hippo	55.8	58.6

Control 2 Hippo	30.6	33.6
Control 4 Hippo	12.2	19.8
Control (Path) 3 Hippo	11.3	11.2
AD 1 Temporal Ctx	21.2	27.3
AD 2 Temporal Ctx	32.3	33.1
AD 3 Temporal Ctx	6.3	10.2
AD 4 Temporal Ctx	24.1	24.3
AD 5 Inf Temporal Ctx	100.0	99.8
AD 5 Sup Temporal Ctx	53.4	55.4
AD 6 Inf Temporal Ctx	51.9	54.9
AD 6 Sup Temporal Ctx	48.3	53.5
Control 1 Temporal Ctx	7.7	11.8
Control 2 Temporal Ctx	31.4	37.9
Control 3 Temporal Ctx	18.0	20.4
Control 3 Temporal Ctx	10.4	12.4
Control (Path) 1 Temporal Ctx	51.1	51.2
Control (Path) 2 Temporal Ctx	28.8	38.7
Control (Path) 3 Temporal Ctx	9.6	11.9
Control (Path) 4 Temporal Ctx	27.6	35.7
AD 1 Occipital Ctx	17.9	22.6
AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 3 Occipital Ctx	10.9	11.6
AD 4 Occipital Ctx	22.1	25.4
AD 5 Occipital Ctx	47.0	36.6
AD 6 Occipital Ctx	32.0	49.3
Control 1 Occipital Ctx	7.4	6.2
Control 2 Occipital Ctx	45.9	48.5
Control 3 Occipital Ctx	15.7	19.5
Control 4 Occipital Ctx	13.6	12.8
Control (Path) 1 Occipital Ctx	96.3	79.6
Control (Path) 2 Occipital Ctx	10.8	13.1
Control (Path) 3 Occipital Ctx	8.5	6.5
Control (Path) 4 Occipital Ctx	13.4	17.7
Control 1 Parietal Ctx	12.0	12.6
Control 2 Parietal Ctx	45.8	58.0
Control 3 Parietal Ctx	14.1	18.6
Control (Path) 1 Parietal Ctx	58.1	70.9
Control (Path) 2 Parietal Ctx	25.9	29.1
Control (Path) 3 Parietal Ctx	12.8	11.0
Control (Path) 4 Parietal Ctx	34.0	45.6

Panel 1.2 Summary: Ag1377 The NOV1 (20936375_0_228_da1) gene is expressed at moderate to high levels in all of the tissues on this panel, with highest expression detected in

a brain cancer cell line (CT=23.3). In general, expression of this gene appears to be higher in cancer cell lines than in normal tissues; this pattern of expression holds true for breast, ovarian and colon cancer cell lines. Thus, therapeutic modulation of the expression or function of this gene, through the use of small molecule drugs, antibodies or protein therapeutics, might be of benefit for the treatment of cancer.

Among tissues derived from the central nervous system, the NOV1 gene is expressed at high levels in fetal brain, amygdala, cerebellum, hippocampus, thalamus, cerebral cortex and spinal cord (CTs=25-27). High expression throughout the brain indicates a potential role for this gene in normal brain function. The NOV1 gene encodes a protein with homology to endozepine. Endogenous benzodiazepine-like substances are thought to play a role in the development of hepatic encephalopathy. Gooday R, Hayes PC, Bzeizi K, O'Carroll RE. Benzodiazepine receptor antagonism improves reaction time in latent hepatic encephalopathy. *Psychopharmacology* (Berl) 1995 Jun;119(3):295-8. It has been suggested that benzodiazepine receptor antagonism may improve cognitive function, particularly speed of information processing, in patients with latent hepatic encephalopathy. Thus, the NOV1 gene product used as a protein therapeutic, or drugs that stimulate the protein's function, may have efficacy in the treatment of hepatic encephalopathy. Increased expression of diazepam binding inhibitor (DBI), a endozepine peptide with anxiogenic action, in Alzheimer's disease, addiction and schizophrenia, indicates that drugs that inhibit NOV1 protein activity may also have utility in the treatment of these diseases. Edgar PF, Schonberger SJ, Dean B, Faull RL, Kydd R, Cooper GJ. A comparative proteome analysis of hippocampal tissue from schizophrenic and Alzheimer's disease individuals. *Mol Psychiatry* 1999 Mar;4(2):173-8. Ohkuma S, Katsura M, Tsujimura A. Alterations in cerebral diazepam binding inhibitor expression in drug dependence: a possible biochemical alteration common to drug dependence. *Life Sci* 2001 Feb 2;68(11):1215-22.

Among tissues with metabolic function, this gene is expressed at high levels in thyroid, adrenal gland, pituitary gland, heart, skeletal muscle, and at moderate levels in pancreas. Therefore, this gene product may be important for the pathogenesis and/or treatment of diseases in any or all of these tissues. In particular, a diazepam binding inhibitor has previously been identified as a potential autoantigen in autoimmune diabetes in a screen of a human pancreatic islet cDNA library with the sera from the diabetic patients autoantigens in autoimmune diabetes. Suk K, Kim YH, Hwang DY, Ihm SH, Yoo HJ, Lee MS. Molecular cloning and expression of a novel human cDNA related to the diazepam binding inhibitor. *Biochim Biophys Acta* 1999 May 31;1454(1):126-31. In addition, treatment of rodents with

the octadecaneuropeptide [diazepam-binding inhibitor (33-50)] has been shown to result in decreased food intake and weight loss. de Mateos-Verchere JG, Leprince J, Tonon MC, Vaudry H, Costentin J. The octadecaneuropeptide [diazepam-binding inhibitor (33-50)] exerts potent anorexigenic effects in rodents. Eur J Pharmacol 2001 Mar 2;414(2-3):225-31. These observations suggest that therapeutic modulation of the diazepam-binding inhibitor-like protein encoded by the NOV1 gene may be useful in the treatment of diabetes and/or obesity.

Panel 1.3D Summary: Ag1865/Ag2029/Ag2813 Results from two of three experiments gave very reproducible results; the results obtained using Ag1865 were somewhat different and will not be discussed here. The expression of this gene is highest in a samples derived from whole brain tissue (CT=27). Among tissues derived from the central nervous system, the NOV1 gene is expressed at high levels in fetal brain, amygdala, cerebellum, hippocampus, thalamus, cerebral cortex and spinal cord (CTs=28-31). Please see Panel 1.2 summary for discussion of potential utility based upon expression in the CNS.

Interestingly, the NOV1 gene is more highly expressed in fetal skeletal muscle (CT=29) than adult skeletal muscle (CT=36-40), suggesting that this gene could be used to distinguish the two. In addition, the increased NOV1 gene expression in fetal skeletal muscle suggests that the protein product may enhance muscular growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the NOV1 gene could be useful in treatment of muscular related disease. More specifically, treatment of weak or dystrophic muscle with the protein encoded by this gene could restore muscle mass or function. Among tissues with metabolic function, the NOV1 gene is expressed at low to moderate levels in pancreas, adrenal gland, pituitary gland, thyroid, heart, liver and adipose. Please see Panel 1.2 summary for discussion of potential utility of this gene in metabolic diseases.

Panel 2D Summary: Ag2813 The NOV1 gene is most highly expressed in a sample derived from normal prostate tissue (CT=25.6). In addition, substantial expression of this gene is detected in a number of tissue samples on this panel. Strikingly, NOV1 gene expression is higher in cancers of the ovary, breast and stomach when compared with their associated normal adjacent tissues. Thus, the expression of this gene could be used to distinguish ovarian, breast or stomach cancer tissue from normal tissue. Moreover, therapeutic modulation of the NOV1 gene expression or activity, through the use of small molecule drugs,

antibodies or protein therapeutics might be of benefit in the treatment of ovarian, breast or stomach cancer.

Panel 4D Summary: Ag1377/Ag1865/Ag2029/Ag2813 Four experiments using three different probe/primer sets gave results that are in good agreement. The NOV1 gene is moderately expressed in the majority of samples on this panel (CT values ranging from 27.5 to 32). However, this gene is expressed at a higher level in activated B cells (Ramos, B cells plus PWM and PBMC plus PWM). For the experiment using Ag1377, the transcript is also observed in macrophages stimulated with LPS. The NOV1 gene encodes a protein with homology to a membrane-associated diazepam binding inhibitor, which has been shown to have immunomodulatory activity. Stepien H, Agro A, Crossley J, Padol I, Richards C, Stanis A. Immunomodulatory properties of diazepam-binding inhibitor: effect on human interleukin-6 secretion, lymphocyte proliferation and natural killer cell activity in vitro. *Neuropeptides* 1993 Sep;25(3):207-11. Therefore small molecules target or antibodies against the NOV1 protein might modulate B cell activity and be useful in the treatment of diseases associated with B cell activation, such as autoimmune diseases (including systemic lupus erythematosus and rheumatoid arthritis) and hyperglobulinemia. In addition, the NOV1 gene is quite abundantly expressed in dermal fibroblasts treated with TNF-alpha, suggesting that therapeutics designed with the protein encoded for by this gene could also be beneficial in the treatment of inflammatory skin diseases such as psoriasis, contact dermatitis, skin infection. Finally, high NOV1 gene expression is seen in the thymus suggesting that this gene product might be involved in the normal homeostasis of the thymus.

CNS_neurodegeneration_panel_v1.0 Summary: Ag1865/Ag2813 Results from two experiments using different probe/primer sets are in good agreement. The NOV1 gene is moderately expressed in all samples in this panel, confirming the expression of this gene in the brain. Gene expression levels are similar in all brain regions tested (hippocampus, temporal cortex, parietal cortex, and occipital cortex) and show no apparent alteration in patients with Alzheimer's disease.

Example 3. SNP analysis of NOVX clones

SeqCalling™ Technology: cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, cell lines, primary cells or tissue cultured primary cells and cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression for example, growth

factors, chemokines, steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled with themselves and with public ESTs using bioinformatics programs to generate CuraGen's human SeqCalling database of SeqCalling assemblies. Each assembly contains one or more overlapping cDNA sequences derived from one or more human samples. Fragments and ESTs were included as components for an assembly when the extent of identity with another component of the assembly was at least 95% over 50 bp. Each assembly can represent a gene and/or its variants such as splice forms and/or single nucleotide polymorphisms (SNPs) and their combinations.

Variant sequences are included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, however, in the case that a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern for example, alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, stability of transcribed message.

Method of novel SNP Identification: SNPs are identified by analyzing sequence assemblies using CuraGen's proprietary SNPTool algorithm. SNPTool identifies variation in assemblies with the following criteria: SNPs are not analyzed within 10 base pairs on both ends of an alignment; Window size (number of bases in a view) is 10; The allowed number of mismatches in a window is 2; Minimum SNP base quality (PHRED score) is 23; Minimum number of changes to score an SNP is 2/assembly position. SNPTool analyzes the assembly and displays SNP positions, associated individual variant sequences in the assembly, the depth of the assembly at that given position, the putative assembly allele frequency, and the SNP sequence variation. Sequence traces are then selected and brought into view for manual validation. The consensus assembly sequence is imported into CuraTools along with variant

sequence changes to identify potential amino acid changes resulting from the SNP sequence variation. Comprehensive SNP data analysis is then exported into the SNPCalling database.

Method of novel SNP Confirmation: SNPs are confirmed employing a validated method know as Pyrosequencing (Pyrosequencing, Westborough, MA). Detailed protocols for Pyrosequencing can be found in: Alderborn et al. Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. (2000). *Genome Research*. 10, Issue 8, August. 1249-1265. In brief, Pyrosequencing is a real time primer extension process of genotyping. This protocol takes double-stranded, biotinylated PCR products from genomic DNA samples and binds them to streptavidin beads. These beads are then denatured producing single stranded bound DNA. SNPs are characterized utilizing a technique based on an indirect bioluminometric assay of pyrophosphate (PPi) that is released from each dNTP upon DNA chain elongation. Following Klenow polymerase-mediated base incorporation, PPi is released and used as a substrate, together with adenosine 5'-phosphosulfate (APS), for ATP sulfurylase, which results in the formation of ATP. Subsequently, the ATP accomplishes the conversion of luciferin to its oxi-derivative by the action of luciferase. The ensuing light output becomes proportional to the number of added bases, up to about four bases. To allow processivity of the method dNTP excess is degraded by apyrase, which is also present in the starting reaction mixture, so that only dNTPs are added to the template during the sequencing. The process has been fully automated and adapted to a 96-well format, which allows rapid screening of large SNP panels. The DNA and protein sequences for the novel single nucleotide polymorphic variants are reported. Variants are reported individually but any combination of all or a select subset of variants are also included. In addition, the positions of the variant bases and the variant amino acid residues are underlined.

Results

Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

NOV1 SNP data:

NOV1 has five SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:1 and 2, respectively. The nucleotide sequences of the NOV1 variants differs as shown in Table 11.

Table 11. cSNP and Coding Variants for NOV1				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change

87	T	G	17	Leu to Arg
545	A	G	170	Thr to Ala
1176	C	T	380	Pro to Leu
1190	C	T	385	Arg to Stop
1246	G	A	403	Met to Ile

Example 4. Molecular cloning of NOV2 variants

For NOV2b, the cDNA coding for the DOMAIN of NOV2a (CG51523-05) from residue 21 to 493 was targeted for “in-frame” cloning by PCR. The PCR template was based on the previously identified plasmid, when available, or on human cDNA(s). For NOV2c-f, the cDNA coding for the DOMAIN of CG51523-05 from residue 43 to 494 was targeted for “in-frame” cloning by PCR. The PCR template was based on human cDNA(s). For NOV2g-r, the cDNA coding for the full-length of CG51523-05 from residue 1 to 523 was targeted for “in-frame” cloning by PCR. The PCR template was based on human cDNA(s).

Table 31. Oligonucleotide primers used to clone the target cDNA sequence:

NOV2 variant	Primers	Sequences
NOV2b	F2	5' -AAGCTTGACAGACCTTGGGACCGGGCCAACTGG-3' (SEQ ID NO: 58)
	R1	5' -CTCGAGAGGAGACATCTCGAAGGGCCACCAAGATGG-3' (SEQ ID NO: 59)
NOV2c-f	F3	5' -AAGCTTACTAGGTTTGAGGCGGCCGTGAAGG-3' (SEQ ID NO: 60)
	R1	5' -CTCGAGAGGAGACATCTCGAAGGGCCACCAAGATGG-3' (SEQ ID NO: 61)
NOV2g-r	F1	5' -AAGCTTCCACCATGTTCCAGTTTCATGCAGGCTCTTGG-3' (SEQ ID NO: 62)
	R2	5' -CTCGAGGTTCACTTTTCTTCTCTTCTTTGATAG-3' (SEQ ID NO: 63)

For downstream cloning purposes, the forward primer includes an in-frame Hind III restriction site and the reverse primer contains an in-frame Xho I restriction site.

Two parallel PCR reactions were set up using a total of 0.5-1.0 ng human pooled cDNAs as template for each reaction. The pool is composed of 5 micrograms of each of the following human tissue cDNAs: adrenal gland, whole brain, amygdala, cerebellum, thalamus, bone marrow, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, liver, lymphoma, Burkitt's Raji cell line, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small Intestine, spleen, stomach, thyroid, trachea, uterus.

When the tissue of expression is known and available, the second PCR was performed using the above primers and 0.5ng-1.0 ng of one of the following human tissue cDNAs:

skeleton muscle, testis, mammary gland, adrenal gland, ovary, colon, normal cerebellum, normal adipose, normal skin, bone marrow, brain amygdala, brain hippocampus, brain substantia nigra, brain thalamus, thyroid, fetal lung, fetal liver, fetal brain, kidney, heart,

spleen, uterus, pituitary gland, lymph node, salivary gland, small intestine, prostate, placenta, spinal cord, peripheral blood, trachea, stomach, pancreas, hypothalamus.

The reaction mixtures contained 2 microliters of each of the primers (original concentration: 5 pmol/ul), 1 microliter of 10mM dNTP (Clontech Laboratories, Palo Alto CA) and 1 microliter of 50xAdvantage-HF 2 polymerase (Clontech Laboratories) in 50 microliter-reaction volume. The following reaction conditions were used:

PCR condition 1:

- a) 96°C 3 minutes
- b) 96°C 30 seconds denaturation
- c) 60°C 30 seconds, primer annealing
- d) 72°C 6 minutes extension

Repeat steps b-d 15 times

- e) 96°C 15 seconds denaturation
- f) 60°C 30 seconds, primer annealing
- g) 72°C 6 minutes extension

Repeat steps e-g 29 times

- e) 72°C 10 minutes final extension

PCR condition 2:

- a) 96°C 3 minutes
- b) 96°C 15 seconds denaturation
- c) 76°C 30 seconds, primer annealing, reducing the temperature by 1 °C per

cycle

- d) 72°C 4 minutes extension

Repeat steps b-d 34 times

- e) 72°C 10 minutes final extension

An amplified product was detected by agarose gel electrophoresis. The fragment was gel-purified and ligated into the pCR2.1 vector (Invitrogen, Carlsbad, CA) following the manufacturer's recommendation. Twelve clones per PCR reaction were picked and sequenced. The inserts were sequenced using vector-specific M13 Forward and M13 Reverse primers and the gene-specific primers in Table 32.

Table 32. Gene-specific Primers

NOV2 variant	Primers	Sequences
NOV2b	SF1	GCAGTCTCTGAAGGACATTCTGCAT (SEQ ID NO: 64)
	SF2	TGTTATTCCTGCCATCTTCTCCTCC (SEQ ID NO: 65)
	SF3	ACTTCACTGTCTGAATCGCTTGTC (SEQ ID NO: 66)
	SF4	TTCTTATGGTCTGCTGACTTCTTCATC (SEQ ID NO: 67)
	SF5	TGACACGCAGCAATTCTTCAACTTT (SEQ ID NO: 68)
	SR1	ACTGCAGAACTGGAAACGCTGACT (SEQ ID NO: 69)
	SR2	GAGAAGATGGCAGGAATAACAGCG (SEQ ID NO: 70)
	SR3	GTTTACTGTGATTCTATGGAACAATTGG (SEQ ID NO: 71)
	SR4	CCATAAGAATTTGGAAGTCATTGTCACTAA (SEQ ID NO: 72)
	SR5	TCATAGGTCCATTTTATGAAATTGTCGAG (SEQ ID NO: 73)
NOV2c-f	SF1	CATGTCCTCCTGCAGTCTCATCA (SEQ ID NO: 74)
	SF2	CTTCAACTGCAACCACCTGCATATTCC (SEQ ID NO: 75)
	SF3	GCTGAATTCCTGTAACATCTTCAACA (SEQ ID NO: 76)
	SF4	TTTTCTTATCATTATCACTTGTTCGCTCC (SEQ ID NO: 77)
	SF5	ATTCTTCAACTTTCTCAGTCATTGGC (SEQ ID NO: 78)
	SR1	GCTGATGAGACTGCAGGAGGACAT (SEQ ID NO: 79)
	SR2	GAAAAGGTGAAGTCAAGCATGGAGG (SEQ ID NO: 80)
	SR3	TCAGCATTGACAAGCGATTCAG (SEQ ID NO: 81)
	SR4	AAGAAGTCAGCAGACCATAAGAATTG (SEQ ID NO: 82)
	SR5	GCTGCGTGTGCATAGGTCCATTTT (SEQ ID NO: 83)
NOV2g-r	SF1	CAAAGCAGTCAGCGTTTCCAGTTTCT (SEQ ID NO: 84)
	SF2	CGCTGTTATTCTCCTGCCATCTTCTC (SEQ ID NO: 85)
	SF3	TCGCTTGTCAAATGCTGAATTCCT (SEQ ID NO: 86)
	SF4	TATCACTTTGTTCTGCTCCTTTCATT (SEQ ID NO: 87)
	SF5	TCAACATATGCAATCATGGCTTCC (SEQ ID NO: 88)
	SR1	GCAAAATCATCAACATCAACATTGCAG (SEQ ID NO: 89)
	SR2	AGGCGGAGAACTGACGAATTCTCTAA (SEQ ID NO: 90)
	SR3	ACAAGCGATTTCAGACAGTGAAGTTTA (SEQ ID NO: 91)
	SR4	TGATAAGAAAATGATGAAGAAGTCAGC (SEQ ID NO: 92)
	SR5	TGAAAAAGATTATTGAAACTATGCCAA (SEQ ID NO: 93)

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims.